THE

INTERNATIONAL SERIES

OF

MONOGRAPHS ON PHYSICS

GENERAL EDITORS

N. F. MOTT, E. C. BULLARD
D. H. WILKINSON

THE INTERNATIONAL SERIES OF MONOGRAPHS ON PHYSICS

Already Published

GEOMAGNETISM

S

BY

SYDNEY CHAPMAN

HIGH ALTITUDE OBSERVATORY, BOULDER, COLORADO
AND
GEOPHYSICAL INSTITUTE, COLLEGE, ALASKA

AND

JULIUS BARTELS

PROFESSOR OF GEOPHYSICS, UNIVERSITY OF GÖTTINGEN

VOLUME I
GEOMAGNETIC AND RELATED
PHENOMENA

Magnus magnes ipse est globus terrestris.
W. GILBERT, *De Magnete*, 1600.

OXFORD
AT THE CLARENDON PRESS

PREFACE

RECENT years have seen the publication of several excellent encyclo-paedia and handbook articles on geomagnetism, an important collective work, and a few brief monographs; but there remains a need, which in this book we have tried to meet, for a comprehensive treatise on the subject. The most recent general expositions of geomagnetism were published in 1900 (Mascart) and 1866 (Walker); they were both limited in scope, and answered few of the questions which most interest modern workers on geomagnetism or on related subjects, like solar and cosmic ray physics, geophysical prospecting, and radio communication.

In 1927 the subject announced for the Adams Prize essay competition, open to graduates of the University of Cambridge, was the theoretical interpretation of the phenomena of the earth's magnetism. As a pre-liminary to the essay for which in 1929 this prize was awarded to one of us (S. C.), an account† of the main facts of the subject was prepared. This, together with the essay itself,‡ was intended to form the basis for a treatise on geomagnetism. In 1929 the co-operation of the second author‖ (J. B.) in the preparation of such a treatise was invited, and since then the writing of the book has been continued jointly—often with long intervals during which other duties or interests intervened. Since the subject has advanced materially in several directions since 1929, our earlier drafts have naturally been much revised and extended.

The wide scope of the subject rendered it desirable to divide the book into three parts, included in two volumes. The first volume, containing Part 1, describes the observed facts of geomagnetism, and how they are measured. It includes also brief accounts of some related phenomena, particularly the leading facts concerning the lunar and solar motions, and the disturbances and other properties of the sun's atmosphere; also concerning earth-currents, the aurora, and the upper atmosphere of the earth. In addition we have included a chapter on magnetic prospecting and the relation of geomagnetism to geology.

The second volume includes Parts 2 and 3 of the book, which deal respectively with the analysis and with the physical interpretation of the phenomena described in Part 1.

The science of geomagnetism has a distinctive character, standing as it does between solar physics and the mainly more local terrestrial

† Dealing with the subjects of Chapters III, VI–IX, XI, XII, XV.
‡ Dealing with the subjects of Chapters XX, XXII–XXV.
‖ To whom Chapters II (B), IV, X, XIII, XIV, XVI–XIX, XXI are mainly due.

science of meteorology, on the one hand, and on the other, the universal science of physics. The discussion of its problems has led to the develop-ment of special techniques, including the invention of graphical, analy-tical, and statistical methods (described in Part 2) for representing and studying the main characteristics, spatial or temporal, of the pheno-mena. These methods can also find fruitful application elsewhere, for example in the vigorous new science of radio physics—which has so much to contribute, with geomagnetism, to our knowledge (now of great practical importance) of the upper atmosphere.

The physical interpretation of geomagnetic phenomena, described in Part 3, is difficult, and this is the part of our book with which we feel least satisfied. Though real progress has been made, some of the chief problems, such as the origin of the main magnetic field of the earth, and the cause of the secular variation, remain unsolved. Part 3 will, we hope, be the first to need revision; but as yet it is not easy to see whether the main advances will come from developments in atomic physics, solar physics, or radio physics.

The last chapter (XXVI) contains a short historical account of the earlier development of geomagnetic science, and at the end of Volume II we give a series of tables of daily magnetic character-figures, daily sunspot numbers, and other data, which we hope will prove convenient to those who study the connexions between geomagnetism and solar or other phenomena.

A special feature of our book is its wealth of illustrations; much care has been taken with the descriptions of these, so as to make them easily understood. The subject of geomagnetism is so intricate and extensive that its progress depends in no small degree on the development of clear and concise methods of representing graphically the multifarious facts found by observation and research; Halley took the first step in this direction by his invention of the isomagnetic world chart.

Another feature of the book which may call for comment is the extent of the bibliography. The study of geomagnetism necessitates reference to a wide range of publications, many of which are not familiar to workers in allied fields, such as astronomy or radio physics. Equally the worker on geomagnetism may not be acquainted with some of the publi-cations which contain contributions to other sciences that bear on his own work. But though the bibliography is so large, many important papers have been omitted, especially in regard to the allied subjects. There, and also in the text, the bias of our personal interests has in-evitably affected our choice of topics and references.

Much labour has been spent on the index of authors and particularly on the subject index. The list of chapter titles (pp. xiii–xxv) suffices to indicate our conception of the main structure of the science of geomagnetism, and the list of section titles indicates the order of the detailed treatment. But many topics are necessarily discussed in several different parts of the book, and the aim of the subject index is to facilitate reference to them, and also to the very numerous details of the subject. Just as in the thousand pages of this book we have tried to collect and set in order the main facts and results of geomagnetism, based on perhaps a hundred thousand pages of printed matter (observations and discussion), so we have endeavoured, by means of the subject index, to make the detailed contents of our thousand pages easily accessible to the reader. Owing to the wide scope of the index, we have not arranged it solely according to the accident of alphabetical order, which would have atomized it into a too extensive medley of disconnected items; instead the use of alphabetical order is restricted to the items (but not necessarily to the sub-items) within each of eleven index-divisions. These divisions have an intrinsic value in showing concisely, though in much detail, the scope and inter-connexions of the main topics.

Throughout the text, the bibliography, and the subject index, care has been taken, by the insertion of chapter and section numbers and other indications at the head of each page, to help the reader in looking up references or cross-references.

S. C.
J. B.

LONDON, *April* 1940
WASHINGTON, D.C., U.S.A., *April* 1940

PREFACE TO THE SECOND EDITION

ALL the publisher's copies of this book were sold by 1948, and the continued demand called for a second edition. After much discussion between the authors, editors, and publishers, it was decided not to rewrite the whole work; this is not necessary at present, and will take much time when the development of the subject makes it imperative.

Instead, the second edition consists of three volumes, of which the first two contain the original text reproduced photographically without correction. The most substantial change in these two volumes is the omission of the Tables (numbered T 1 to T 77 in the original edition) from Volume II. Volume III contains: (a) a series of supplements to the original chapters, describing the progress of the subject since the first edition was written; (b) a supplementary bibliography; (c) an improved and extended set of Tables; (d) a list of corrections to Volumes I and II; (e) a new author's index; and (f) a new subject-index. The latter index is not subdivided as in the original edition, because it appears that many readers found the division confusing rather than helpful.

It has been thought useful, however, to reproduce in Volume II of this edition the original author's index and the (divided) subject-index relating to Volumes I and II.

<div style="text-align: right">

S. C.

J. B.

</div>

PREFACE TO THE THIRD IMPRESSION

WITH the advent of Space Science, much of the contents of this book has attained a new interest. It has therefore been considered worth while to reprint the book just as in the second edition, with a few dozen corrections of mistakes. The good intentions of preparing a third volume as expressed in the preface to the second edition still linger; quite a number of tables have already been prepared for the printer.

<div style="text-align: right">

S. C.

</div>

Geophysical Institute, College, Alaska
and High Altitude Observatory, Boulder, Colorado

<div style="text-align: right">

J. B.

</div>

Geophysikalisches Institut, Universität, Göttingen,
and Institut für Stratosphärenphysik of the Max
Planck-Institut für Aeronomie, Lindau bei Göttigen

September 1961

ACKNOWLEDGEMENTS

CONTENTS

b

VOLUME II

PART 2. THE ANALYSIS AND SYNTHESIS OF GEOMAGNETIC DATA

Part 3. PHYSICAL THEORIES OF GEOMAGNETIC PHENOMENA

LIST OF PLATES

VOLUME I

VOLUME II

NOTE REGARDING REFERENCES

In this book the chapter-sections, equations, and figures (but not the plates) are numbered separately in each chapter. References to sections, equations, and figures are made as in the following examples:

	Section	Equation	Figure
In the same chapter	§ 4	(4)	Fig. 4
In another chapter (e.g., IX)	9.4	9 (4)	Fig. 9.4

References to books or papers listed in the bibliography at the end of Volume II are enclosed in square brackets, e.g. [17] if the reference is to the bibliography for the current chapter, and [9.17] if the reference is to the bibliography for some other chapter, such as IX; references to the General Bibliography, which precedes those for the separate chapters, are distinguished by the insertion of the letter G, as in [G 9].

.

PART 1

THE OBSERVED PHENOMENA OF GEOMAGNETISM

THE general plan of this book is to give in Part 1 an account of the observed facts of geomagnetism, and the methods by which they are found and recorded; and to indicate in Part 2 how the great array of data described in Part 1 are analysed and synthesized, as a step towards the discussion, in Part 3, of the physical causes and mechanism of the phenomena.

This general plan is followed broadly, but not strictly, in Part 1. Chapter I includes a brief description of the principles of electromagnetism, so far as it is needed in the later chapters; and in Chapter II, the more detailed second part (B) of the account of measuring instruments and methods refers to spherical harmonic functions, which are described in Part 2 (Chapter XVII); this part of Chapter II may be omitted by readers not specially interested in instrumental questions. Part 1 also includes four chapters describing data other than, but closely concerned with, those of geomagnetism, namely solar and lunar data (Chapter V), earth-currents (Chapter XIII), aurorae (Chapter XIV), and the upper atmosphere (Chapter XV); these chapters include some theoretical topics as well as a description of observed phenomena.

The strictly geomagnetic chapters of Part 1 deal first with the principles and methods of measurement, and include an account of international co-operation in observing the earth's field—a matter of great importance in this as in other branches of geophysics. After a discussion, in Chapters III and IV, of the main magnetic field, its secular variation, and the local anomalies which are of geological interest and of great importance in magnetic prospecting, the remainder of Part 1 is concerned with the transient magnetic variations, periodic and irregular. A general review of these is given in Chapter VI, as the three main types of transient variation are so closely interwoven in the data that it is scarcely possible to describe any one in detail without having at least briefly described the others. Chapters VII and VIII discuss the two main periodic magnetic variations, which are daily, and due to the sun and moon respectively.

Chapters IX–XII deal with various aspects of the irregular variations, or *disturbance*, of the geomagnetic field. Chapter IX relates to the morphology, or geographical distribution and time-variation, of the main type of magnetic disturbances, which in their more extreme forms

are called magnetic storms; a disturbance of this type can be regarded as a unitary phenomenon, with a more or less clearly marked beginning and a fairly standard mode of progress and decay. Some minor types of disturbance are considered in Chapter X. Chapter XI deals with various statistical features, and solar relationships, of magnetic disturbance, and Chapter XII describes the notable 27-day recurrence tendency due to the rotation of the sun, and the consequent changes of presentation of disturbed solar regions towards the earth.

I

GENERAL PRINCIPLES

1.1. Magnetic elements. A magnet free to turn about its mass-centre takes up a definite direction at any point near the earth, thus indicating the presence of magnetic forces. The intensity and direction of the force vary from point to point and from time to time. A region in which a distribution of magnetic force can be detected is called a *magnetic field*.

(Accurate direct measurements of the force are confined to points close to the earth's surface; they have not been made at any great height above the ground, nor within the earth except at relatively shallow depths.) With this limitation, the present general distribution of magnetic force over the earth is known with considerable detail and accuracy. This has been achieved as the result of many series of measurements (called *magnetic surveys*) by sea and land. The time-variations of the field are continuously recorded at a number of *magnetic observatories* (2.2).

The magnetic force at any point can be specified in several ways: for example, by means of the rectangular components X, Y, Z, defined as follows: X is the component along the horizontal direction in the geographical meridian, and is reckoned positive if northward, negative if southward; Y is the horizontal component transverse to the geographical meridian, and is reckoned positive if eastward, negative if westward; Z is the vertical component, reckoned positive if downward, negative if upward. The vertical component Z is often called the *vertical intensity* or *vertical force* (V.F.); the symbol V may be used to denote its numerical magnitude, without regard to sign, so that $V = Z$ where Z is positive, $V = -Z$ where Z is negative; but in this book V will be used in a different sense, to denote the magnetic potential (§ 4).

Another common way of specifying the magnetic force is by means of H, D, and I, defined as follows: H is the magnitude of the horizontal component, considered positive whatever its direction; it is called the *horizontal intensity* or *horizontal force* (H.F.); D is the azimuth of the horizontal component, reckoned positively from the geographical north towards the east, from 0° to 360°, or negatively westwards from the north; it is called the *magnetic declination* or, by mariners, the *magnetic variation* of the compass; I is the angle made by the direction of the whole magnetic force with the horizontal, and is reckoned positive if

the force is inclined downwards, or negative if upwards; it is called the *magnetic dip* or *magnetic inclination*; D and I are measured in degrees and minutes of arc.

The intensity of the whole magnetic force will be denoted by F; it is called the *total intensity* or *total force.*

The magnetic force and its horizontal component, regarded as vectors, having direction as well as magnitude, will be denoted by F and H.
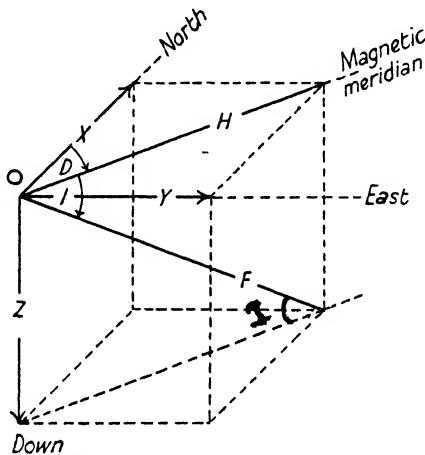
Fig. 1. The geomagnetic force F, its rectangular components X, Y, Z, and the elements H, D, I

The quantities denoted by X, Y, Z, D, I, H, F are called *magnetic elements*. They are evidently connected by the following relations (cf. Fig. 1):

$$\left.\begin{array}{ll} H = F\cos I, & Z = F\sin I = H\tan I, \\ X = H\cos D, & Y = H\sin D, \\ X^2+Y^2 = H^2, & X^2+Y^2+Z^2 = H^2+Z^2 = F^2. \end{array}\right\} \qquad (1)$$

Three independent elements are required in specifying F, and when three such are given, any of the other elements can be determined from the relations (1).

In this book the term magnetic variation will not be used (as by seamen) for the declination D, but will always signify the time-changes in the magnetic field (§ 6).

The vertical plane through the magnetic force (or its horizontal component) is called the *local magnetic meridian*; it makes the angle D with the geographical or astronomical meridian.

The ordinary compass-needle is weighted so as to swing horizontally around a vertical axis. A perfectly balanced magnet, free to swing in

the magnetic meridian plane around a horizontal axis that is perpendicular to this plane, is called a *dip-needle*. Over most of the northern hemisphere the north-seeking end of such a needle will dip downwards (I positive); over most of the southern hemisphere the south-seeking end is the lower one (I negative). The regions of positive and negative I are separated by a curve along which $I = 0$; an unweighted (perfectly balanced) magnetic needle is horizontal at points along this curve. The curve is called the *magnetic equator* or *dip-equator*.

At points where the horizontal component of F vanishes, the dip-needle rests with its axis vertical. Such a point is called a *pole of magnetic dip* or *dip-pole* (see p. 104). There are two principal poles of this kind, often called *the* magnetic poles, one near the north and the other near the south *geographic* poles (i.e. at a distance of a few degrees of latitude); in addition there are some local dip-poles due to purely local irregularities of the earth's field.

1.2. Intrinsic and relative elements. At any point P on a sphere like the earth there is a natural direction characteristic of the point, namely, the radial direction. Since F, H, Z, and I refer either to this direction, or to none at all, they may be called *intrinsic* magnetic elements [8]. But it is not possible to specify F numerically by these elements alone. In order to indicate the direction of the horizontal component, some direction from which to measure azimuth must be chosen. On the earth it is found convenient to adopt the geographic meridians, passing through the axis of the earth's rotation, as the reference lines for azimuth, and the magnetic declination D is measured relative to them. But the axis of rotation has no obviously necessary relation to the earth's magnetic field, so that D, and likewise X and Y, involve reference to something which is extraneous to the field, and chosen for convenience. Thus they may be called *relative* magnetic elements.

1.3. Magnetic intensity units. The unit of magnetic intensity is most simply defined in terms of magnetic poles. These are of two kinds: those of like kind repel each other, whereas unlike poles attract one another. One kind is, by convention, called positive, and the other negative; alternatively they are called 'north-seeking' and 'south-seeking'; these terms are often shortened to 'north' and 'south'. Although magnetic poles are nothing but theoretical conceptions, they are useful for describing the interaction of fields and magnets. A pole is said to be of unit strength if it exerts unit mechanical force on an equal pole placed at unit distance from it; in the c.g.s. system (now

universally used in geomagnetic work) the units of force and distance
are the dyne and the centimetre. A pole of strength m is one which at
unit distance exerts a force m on a unit pole; at distance r the force
on a unit pole is m/r^2, and that on a pole of strength m_1 is mm_1/r^2
(Coulomb's law).

In any magnetic field the mechanical force exerted at a point P on
a pole of strength m_1 is proportional to m_1, so long as the presence of
the pole does not appreciably modify the magnetic state of the bodies
producing the field: this is ensured if m_1 is small; the mechanical force,
divided by m_1, is called the *magnetic intensity* at P. The c.g.s. unit of
magnetic intensity is called a *gauss* (see, however, § 17); it is denoted
by 1 Γ; its physical 'dimensions' are cm.$^{-\frac{1}{2}}$ gm.$^{\frac{1}{2}}$ sec.$^{-1}$ = dyne$^{\frac{1}{2}}$ cm.$^{-1}$
In geomagnetism a smaller unit, called the gamma (γ), is often used;
1 $\Gamma = 10^5 \gamma$. (The unit γ was introduced by Eschenhagen [7] in 1896,
and is now internationally adopted.)

Since geomagnetic investigations often require reference to observa-
tory records and other papers published many years ago, it is desirable
to mention some units formerly adopted. Gauss and Weber used the
mm.$^{-\frac{1}{2}}$ milligram$^{\frac{1}{2}}$ sec.$^{-1}$ unit (0·1 Γ), and formerly this was employed
in several magnetic publications; many British writers and publications
have used the foot$^{-\frac{1}{2}}$ grain$^{\frac{1}{2}}$ sec.$^{-1}$ unit, equal to 0·046108 Γ, so that
1 Γ = 21·688 of these 'British units'. Before Gauss [ref. 1, Ch. II] intro-
duced his 'absolute' method of determining the magnetic intensity, in
terms of the units of length, mass, and time, measurements of the
relative intensity at any station P were made by comparing the times
of oscillation (§ 11) of a particular compass needle at P and at a standard
station, at which F was taken as unity; Humboldt, who made such
observations, chose Micuipampa, Peru, on.the magnetic equator, as his
standard. The unit of intensity thus defined was used at first by Gauss
in his general theory of the earth's magnetic field; afterwards he esti-
mated its value as 0·3494 Γ.

At Greenwich and other magnetic observatories the *variations* of H
and Z were formerly expressed as fractions of H and Z, so that as these
elements changed in magnitude from year to year with the secular
variation, the unit employed for the recorded variations gradually
altered.

1.4. The potential due to a magnetic pole. The magnetic in-
tensity F_1 at a point P due to a pole of strength m_1 at a point P_1 is
m_1/ρ_1^2, in the direction $P_1 P$ if m_1 is positive; here ρ_1 denotes the distance
$P_1 P$.

If the rectangular coordinates of P and P_1 relative to axes through an origin O are x, y, z and x_1, y_1, z_1,

$$\rho_1^2 = (x_1-x)^2+(y_1-y)^2+(z_1-z)^2. \tag{1a}$$

We may regard x, y, z and x_1, y_1, z_1 as the components of vectors r and r_1, equal to OP and OP_1. These vectors are called the *position vectors* of P and P_1. The corresponding components of the magnetic intensity F_1 at P due to m_1 at P_1 are X_1, Y_1, Z_1, where

$$F_1 = m_1/\rho_1^2,$$

$$X_1 = \frac{m_1(x-x_1)}{\rho_1^3}, \qquad Y_1 = \frac{m_1(y-y_1)}{\rho_1^3}, \qquad Z_1 = \frac{m_1(z-z_1)}{\rho_1^3}. \left.\right\} \tag{2}$$

These equations may also be expressed in the form

$$X_1 = -\frac{\partial V_1}{\partial x}, \qquad Y_1 = -\frac{\partial V_1}{\partial y}, \qquad Z_1 = -\frac{\partial V_1}{\partial z}, \tag{3}$$

where

$$V_1 \equiv m_1/\rho_1; \tag{4}$$

V_1 is called the *magnetic potential* at P due to the pole m_1 at P_1, and the intensity F_1 is said to be the *negative gradient* of this potential; in vector notation the equations (2) may be summed up in the one vector equation

$$F_1 = -\operatorname{grad} V_1, \tag{5}$$

since $\operatorname{grad} V_1$ denotes the vector whose components are $\partial V_1/\partial x$, $\partial V_1/\partial y$, $\partial V_1/\partial z$.

If P′ is a point near P, with coordinates $x+\delta x$, $y+\delta y$, $z+\delta z$, and position vector $r+\delta r$ (so that δr is a small vector with components δx, δy, δz) and if $V_1+\delta V_1$ is the potential at P′, then

$$\delta V_1 = \frac{\partial V_1}{\partial x}\delta x+\frac{\partial V_1}{\partial y}\delta y+\frac{\partial V_1}{\partial z}\delta z,$$

which in vector notation can be expressed in the form

$$\delta V_1 = \operatorname{grad} V_1 \cdot \delta r,$$

where the dot is used as the sign for the scalar product of two vectors. By (5) this is equivalent to

$$\delta V_1 = -F_1 \cdot \delta r = -F_1 \,\delta r \cos\theta,$$

where δr denotes the magnitude of δr (that is, the length PP′), and θ denotes the angle between F_1 and δr. Surfaces at all points of which the potential V_1 has the same value are called *equipotential surfaces*. If P′ lies on the equipotential surface through P, so that $\delta V_1 = 0$, it follows from the last equation that $\cos\theta = 0$ or $\theta = 90°$ (since $F_1 \neq 0$, $\delta r \neq 0$). Hence F_1 is perpendicular to all displacements PP′ from P along the

equipotential surface through P, that is, F_1 is normal to this surface, at P.

Evidently, for a given length of displacement δr, δV_1 is greatest when $\cos\theta = -1$, that is, when the direction of δr is opposite to that of F_1; consequently F_1 lies along the direction in which V_1 decreases most rapidly from P.

The work done by the magnetic force F_1 which acts on a positive unit pole placed at P, when the pole is allowed to move to a neighbouring point P', is $F_1 \delta r \cos\theta$ or $-\delta V_1$. The work done when the unit pole is moved from P to *any* other point P', at whatever distance from P, is $-\int dV_1$, integrated along the path from P to P'; this is $-(V_1'-V_1)$, where V_1' is the potential at P'. The work done thus depends only on the positions of the end-points of the path, that is, upon the values of V_1 there, and is independent of the path taken in moving from P to P'. If P' is at infinity, $V_1' = 0$, and the work done is V_1. Consequently V_1 is the work done *by* the magnetic force (or, as we may say, by the field) in moving the unit pole from the point P to infinity. It may also be regarded as the work which must be done by external forces *against* the field, to bring the unit pole from infinity to P.

1.5. Compound magnetic fields. The total magnetic intensity F at P, with components X, Y, Z, due to any distribution of magnetic poles of strengths m_1, m_2,... at points P_1, P_2,... at distances ρ_1, ρ_2,... from P, is given, as in § 4, by

$$X = X_1+X_2+..., \qquad Y = Y_1+Y_2+..., \qquad Z = Z_1+Z_2+..., \quad (6)$$

or, in vector notation,

$$F = F_1+F_2+...;$$

that is, the field is simply the sum of the separate fields. Evidently, therefore,

$$X = -\partial V/\partial x, \quad Y = -\partial V/\partial y, \quad Z = -\partial V/\partial z, \quad F = -\mathrm{grad}\,V,$$

where $\qquad V = V_1+V_2+... = (m_1/\rho_1)+(m_2/\rho_2)+....$ $\qquad\qquad$ (7)

It is often convenient to refer to the whole of a magnetic field, specified by the intensity F given as a function of position and time, by a single symbol; F will generally be used for this purpose, particularly in the case of the earth's magnetic field.

Any magnetic field F will be called the sum of two fields F' and F", specified by their distributions of magnetic intensity F' and F'' in space and time, if at every point and instant $F = F'+F''$, so that also

$X = X' + X''$, etc., as above. The relation between F, F', and F" may be expressed symbolically by

$$F = F' + F''. \tag{8}$$

The notation $-F$ will be used to signify a field in which at every point and instant the intensity is $-F$. The 'difference' between two fields F' and F" will signify the field whose intensity is $F' - F''$; it will be denoted by F' − F". This usage can obviously be extended to the 'addition' or 'subtraction' of any number of magnetic fields, and the operations clearly obey the commutative and associative laws.

When $F = F' + F'' + ...$, the fields F', F",... may be called 'parts' of the field F. Any field F can be analysed into parts in any manner, but the analysis is usually of interest only if the parts F', F",... arise from different causes, or if they vary with time, or are distributed in space, in some physically or mathematically significant way. In free space the magnetic intensity F is the negative gradient of a potential V of the form $\sum (m'/\rho')$, as in § 4, and in any physically significant analysis of F the component fields in free space must each have the same property, being negative gradients of potentials V', V'',..., each of which is the sum of elements m_1/ρ_1.

It may be noted that the potential V and the elements X, Y, Z are additive, whereas in a compound field made up of parts F', F",... this is not in general true of the elements H, F, D, and I.|

1.6. Mean fields and variation fields. If a field F is variable with the time, it is often convenient, when investigating its changes during any interval i, to divide it into two parts, one of which, denoted by $\overline{F_i}$, is equal to the mean value of F during the interval, while the other, denoted by $\Delta_i F$, is the difference $F - \overline{F_i}$. The various elements of $\overline{F_i}$ and $\Delta_i F$ may be denoted in a similar way, so that, for example; at any point P and time t during the interval,

$$\left. \begin{array}{ll} \overline{F_i} = \dfrac{1}{i} \int_i F \, dt, & \Delta_i F = F - \overline{F_i}, \\[3mm] \overline{X_i} = \dfrac{1}{i} \int_i X \, dt, & \Delta_i X = X - \overline{X_i}. \end{array} \right\} \tag{9}$$

If the intensity F of the field F is the negative gradient of a potential V, F_i will be similarly related to a potential $\overline{V_i}$, given by $V_i = \dfrac{1}{i} \int_i V \, dt$, and $\Delta_i F$ will be derived from the potential $\Delta_i V$ given by $\Delta_i V = V - \overline{V_i}$. The field $\overline{F_i}$ and its potential $\overline{V_i}$ are functions of position only, whereas

$\Delta_i \mathsf{F}$ and $\Delta_i V$ are also functions of the time. The two fields may be called the mean and the variation fields for the interval i. The suffix i is usually omitted when the interval concerned is clearly understood.

In the case of the earth's field, X, Y, and Z are known only at a limited number of points, as functions of the time; the mean and variation fields during any interval are determinable only at the same points.

If ΔF is small compared with the mean value $\overline{F}$, the relations (1) may be differentiated as follows:

$$
\left.
\begin{aligned}
\Delta H &= \Delta F \cos I - \Delta I\, F \sin I, \quad & \Delta X &= \Delta H \cos D - \Delta D\, H \sin D, \\
\Delta Z &= \Delta F \sin I + \Delta I\, F \cos I, \quad & \Delta Y &= \Delta H \sin D + \Delta D\, H \cos D, \\
\Delta Z &= \Delta H \tan I + \Delta I\, H / \cos^2 I.
\end{aligned}
\right\} \quad (10)
$$

For small changes, ΔH and $H\,\Delta D$ are horizontal rectangular field components measured in the direction of H and perpendicular to it. If $(\Delta D)'$ is the value of ΔD expressed in minutes,

$$\Delta D = (\Delta D)'/3438; \text{ similarly } \Delta I = (\Delta I)'/3438, \quad (11)$$

the number 3,438 being one radian expressed in minutes of arc.

The symbols X, Y, Z, H, ΔX, ΔY, ΔZ, and V for the potential (not for the vertical force), were adopted by the International Meteorological Conference (2.17) at Paris, 1896.

1.7. Laplace's equation. By (1 a), p. 5,

$$\frac{\partial^2}{\partial x^2}\left(\frac{1}{\rho_1}\right) = \frac{3(x-x_1)^2-\rho_1^2}{\rho_1^5}, \quad \text{etc.,} \quad (12)$$

so that

$$\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}\right)\frac{1}{\rho_1} = 0. \quad (13)$$

The differential operator on the left is often denoted by ∇^2, so that (13) may be written in the form

$$\nabla^2\frac{1}{\rho_1} = 0.$$

Hence any potential function V which is the sum (or integral) of terms m_1/ρ_1 satisfies the equation

$$\nabla^2 V = 0, \quad (14)$$

which is called Laplace's equation. This is satisfied, at points in free space, by the potential of any magnetic field, however produced, because such a field can always be regarded as due to a distribution of magnetic poles.

1.8. Equipotential surfaces and lines of force. Surfaces over which V has a constant value are called *equipotential surfaces*. In a

plane diagram they may be illustrated by the *equipotential curves* in which they are cut by the plane of the figure.

Lines drawn so that, at each point, they have the same direction as the magnetic intensity there are called *lines of (magnetic) force*; it may
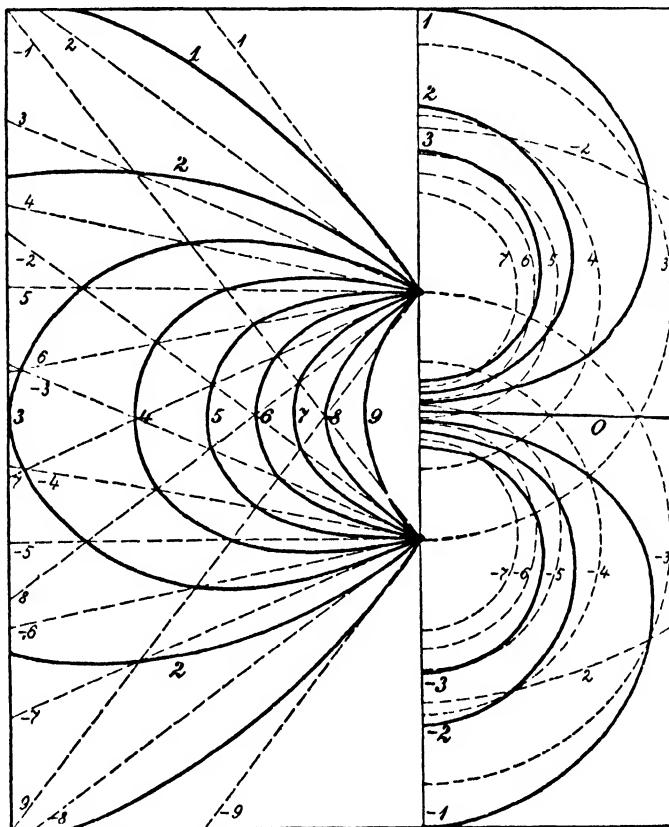
FIG. 2. Schematic magnet (a pair of poles): lines of force (*left*) and sections of equipotential surfaces (*right*). The dotted lines are the lines of force and the sections of the equipotential surfaces for the separate poles; from these, by a method due to Maxwell, the lines for the combined field can be drawn, through the intersections of the lines for the separate fields

be shown, as in § 4, that these lines are everywhere perpendicular to the equipotential surfaces. In general they are not plane curves.

Lines drawn on the earth's surface, in the direction of the horizontal intensity at each point, are called lines of horizontal (magnetic) force or *magnetic meridians*.

In the field of a single magnetic pole, the equipotential surfaces are spheres about the pole as centre, and the lines of force are radial

straight lines diverging from the poles. Fig. 2 illustrates the equipotential surfaces and lines of force for a so-called *schematic magnet* consisting of a pair of equal and opposite magnetic poles. It shows also how the equipotential curves can be constructed by graphical superposition of the individual fields of the two poles [2] (whose equipotential curves are shown by dotted lines), by the method due to Maxwell and described in his treatise *Electricity and Magnetism* (§ 123).

1.9. The 'elementary' magnet or dipole. In any limited portion of matter positive and negative magnetic poles always occur in conjunction, of equal total strengths, so that the net strength is zero. The simplest combination consists of an *'elementary'* magnet or *dipole* or *doublet*, comprising a pair of opposite poles of strengths $-m$ and $+m$, at points P_-, P_+ at a distance d apart, d being infinitesimal and m correspondingly large so that md has a finite value M; M is called the *magnetic moment* of the dipole, and the line $P_- P_+$ is called the *axis*. The axis is considered as having the direction $P_- P_+$ in this 'sense'; interchange of the two poles would therefore reverse the axis. The position of the magnet is specified by that of its centre O, to which P_- and P_+ are infinitesimally close; the moment can be specified by the vector M, of magnitude M and direction $P_- P_+$. If the vector $P_- P_+$ is denoted by d, then $M = md$.
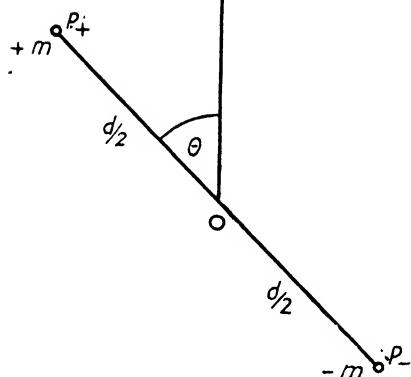
Fig. 3

The position of any other point P relative to the magnet can be specified (Fig. 3) by the position vector $\underline{\text{O}}\text{P}$ or r, of length r and inclined at the angle θ to the axis $P_- \text{O} P_+$. Then, neglecting d^2, the distances $P_- P$ and $P_+ P$ are $r + \frac{1}{2}d\cos\theta$ and $r - \frac{1}{2}d\cos\theta$, so that the potential at P, by (4) and (7), is given by

$$V = \frac{m}{r - \frac{1}{2}d\cos\theta} + \frac{-m}{r + \frac{1}{2}d\cos\theta} = \frac{M\cos\theta}{r^2}. \tag{15}$$

The magnetic intensity at P is clearly in the plane through P and the axis; its 'radial' and 'tangential' components (denoted by R, S), that

is, its components along and perpendicular to OP, in the direction of increasing r and θ, are given by

$$R = -\frac{\partial V}{\partial r} = \frac{2M\cos\theta}{r^3}, \qquad S = -\frac{\partial V}{r\,\partial\theta} = \frac{M\sin\theta}{r^3}. \qquad (16)$$

The resultant intensity $\bar{F}$ at P consequently makes an angle α with OP, given by

$$\tan\alpha = S/R = \tfrac{1}{2}\tan\theta; \qquad (16\,a)$$

this angle is independent of r, showing that the lines of force cut OP at the same angle along its whole length.

Owing to the factor $1/r^3$ in R and S, the field-intensity due to a magnet decreases very rapidly with increasing distance from the magnet.

From (16) it is easily seen that the physical 'dimensions' of a magnetic moment M are gauss $\times$ cm.^{3}

The equation for the lines of force is derived from the condition that the radial and transverse components of an element of the line, dr and $r\,d\theta$, stand in the ratio of the force components, i.e.

$$\frac{dr}{r\,d\theta} = \frac{R}{S} = 2\cot\theta,$$

so that $\qquad \dfrac{dr}{r} = 2\cot\theta\,d\theta = \dfrac{2\sin\theta\cos\theta\,d\theta}{\sin^2\theta} = \dfrac{d(\sin^2\theta)}{\sin^2\theta};$

on integration this gives

$$\log r = \log\sin^2\theta + \text{constant},$$

or $\qquad\qquad\qquad \dfrac{r}{\sin^2\theta} = \text{constant}. \qquad (17)$

1.10. The mechanical couple and force on a dipole in a magnetic field. Consider an elementary magnet situated at a point O in a *uniform magnetic field* F; 'uniform' signifies that the direction and intensity of the field, that is, of F, are everywhere the same. Let the axis of the magnet make an angle θ with the direction of F. Then the poles $-m$, $+m$ of the magnet are acted on by equal and opposite mechanical forces $\pm mF$; hence there is no resultant mechanical force on the magnet. The forces act, however, along lines at a distance $d\sin\theta$ apart; hence they constitute a mechanical couple of moment $mF\times d\sin\theta$ or $MF\sin\theta$; this couple tends to turn the magnet about its centre, towards alinement with the field (Fig. 4a). In the notation of vector analysis, this couple can be expressed (using the sign $\wedge$ for vector product) as $M\wedge F$, which indicates the direction of the axis about which the

couple tends to turn the dipole, and the sense of the moment, as well as the magnitude of the couple.

If the field F is not uniform, however, the intensities at P_- and P_+ are not quite identical. They will differ from the intensity F at O by small amounts which (since d is small and ultimately zero) will be equal and opposite. Let $F+\delta F$ denote the intensity at P_+; then that at P_-
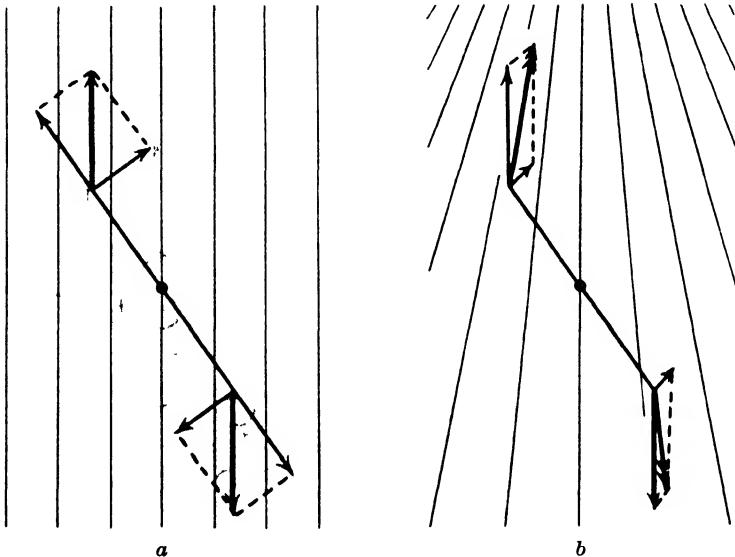
FIG. 4. Forces acting on a schematic magnet (a pair of poles), in a uniform field (*left*) and in a non-uniform field (*right*): showing the lines of force, the couple, and the resultant force

is $F-\delta F$. The forces on the two poles are therefore $m(F+\delta F)$ and $-m(F-\delta F)$. The resultant mechanical force is now $2m\,\delta F$, with components $2m\,\delta X$, $2m\,\delta Y$, $2m\,\delta Z$. Each of the components X, Y, Z is a scalar function of position; hence δX, the change of X from O to P_+, is $\frac{1}{2}d\cdot\operatorname{grad}X$ (cf. § 4) or $\frac{1}{2}\left(d_x\frac{\partial X}{\partial x}+d_y\frac{\partial X}{\partial y}+d_z\frac{\partial X}{\partial z}\right)$, where d_x, d_y, d_z denote the x-, y-, z-components of d; and the x-component of the mechanical force on the magnet becomes $2m\,\delta X$ or $md\cdot\operatorname{grad}X$ or $M\cdot\operatorname{grad}X$. For instance, the mechanical force exerted on a dipole of moment M_1 in the field of another parallel dipole of moment M_2, situated, at distance r, in a direction perpendicular to the axes of *both* the dipoles, is $3M_1M_2/r^4$; it is repulsive if the axes of the dipoles are parallel and have the same sense, and attractive if they are anti-parallel.

The moment of the forces on the poles, about the centre of the dipole,

is the same as the couple in the case of the uniform field, namely, $M \wedge F$, since the sum of the turning moments of the two forces is $\frac{1}{2}d \wedge m(F+\delta F)+(-\frac{1}{2}d) \wedge \{-m(F-\delta F)\}$; this differs from $M \wedge F$ only by terms which vanish with d.

From (16) it follows that, in a field of a dipole, the couple on another dipole decreases, with increasing distance, proportionately to $1/r^3$; the resultant force decreases more rapidly, like $1/r^4$. The effect of the inhomogeneity of the field in producing both a couple and a resultant force is shown schematically in Fig. 4 b.

In measuring magnetic fields, only the couple is used.

Bar-magnets of finite size can for many purposes be regarded as approximately equivalent to a *schematic magnet*, i.e. to a pair of poles of finite strengths $-m$, $+m$ situated at points near the ends of the magnet, the distance d between them being about 0·8 of the whole length l of the bar. Their magnetic moment M may be defined as the quotient of the couple experienced when placed in a field of uniform intensity H, at a small inclination θ to the field, divided by $H \sin \theta$.

In terms of this value of M, the formulae (16) for R and S apply approximately to the field of a bar-magnet; experience shows that their error does not exceed 1 per cent. at distances beyond about $5l$. The accurate potential V contains, however, terms additional to $M \cos \theta/r^2$, which involve higher powers of $1/r$. They contribute appreciably to the intensity at points near to the magnet, and must be taken into account in accurate measurements (Ch. II).

1.11. The oscillation of a magnet in a magnetic field. If a horizontal elementary magnet (pole strength m, distance d, moment $M = md$) is free to turn about a vertical axis in a horizontal field of intensity H, it can be at rest only when alined along the magnetic meridian. If turned out of this position by a small angle θ (in circular measure) it becomes subject to the restoring couple $2(\frac{1}{2}d)mH \sin \theta$ or $MH \sin \theta$ or, approximately, $MH\theta$, since θ is small. Hence its angular acceleration $\ddot{\theta}$ ($\equiv d^2\theta/dt^2$, $t =$ time) towards its equilibrium direction is given by the equation

$$I\ddot{\theta} = -MH\theta, \tag{18}$$

where I denotes the mechanical moment of inertia of the magnet about the vertical line through the point of support O. The support may be afforded by a pivot, or (as is usual in actual observations) by a suspending thread of negligible torsion. The solution of (18) is

$$\theta = \theta_0 \sin(nt+\epsilon), \tag{19}$$

where θ_0 and ϵ are arbitrary constants, and

$$n^2 = MH/\mathrm{I}. \tag{20}$$

Thus the magnet performs simple harmonic oscillations with the period T given by

$$T = 2\pi/n, \qquad T^2 = 4\pi^2\mathrm{I}/MH. \tag{21}$$

Similarly, the magnet can oscillate, about a horizontal axis, in the vertical plane containing the direction of an inclined field of intensity F, with the period $2\pi\sqrt{(\mathrm{I}/MF)}$.

The mathematical analogy with the oscillations of a simple or compound pendulum in the gravitational field of the earth is obvious. If the oscillations have a finite range $2h$ (reckoned in circular measure), the observed value of the period T is corrected to the value corresponding to 'infinitely small oscillations' by multiplication by $(1-h^2/16)$; for instance, by $0\cdot999981$ if h is $1°$, which in circular measure is $0\cdot01745$.
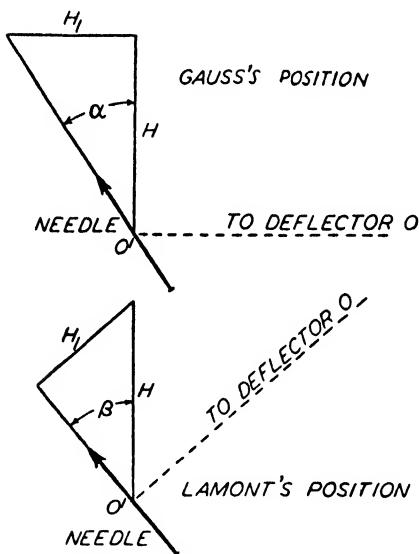
1.12. The deflexion of a magnet by another magnet. If a small magnetized needle is at rest at a point O', in a horizontal magnetic field of intensity H, and another magnet (the *deflector*) of moment M is brought into the field, with its centre at a point O distant r from O', and with its axis in the direction OO', it will modify the field at O by the addition of a magnetic intensity H_1 in the direction OO'. The value of H_1 is $2M/r^3$ (by (16), putting $\theta = 0$), if the deflector is very small, or $2kM/r^3$, where k is nearly unity, if the length of the deflector is not very small compared with r; k is called the *deviation constant* or *deflexion constant* of the magnet and will depend on r; it approaches unity with increasing r (cf. § 10, last paragraph).

If OO' is perpendicular to the field H, the needle will take up a new position inclined at the angle α to the original one, where

$$\tan\alpha = H_1/H = 2kM/Hr^3. \tag{22}$$

But if OO' is chosen so that it is perpendicular to the new position of

GAUSS'S POSITION

NEEDLE

TO DEFLECTOR O

TO DEFLECTOR O

LAMONT'S POSITION

NEEDLE

FIG. 5

equilibrium of the needle in the combined field, the deflexion will be β, given by

$$\sin \beta = H_1/H = 2kM/Hr^3. \tag{23}$$

The two cases are often referred to as *Gauss's* and *Lamont's positions* (Fig. 5).

1.13. The magnetic field due to an electric current-circuit.

Oersted, in 1819, discovered that an electric current flowing in a circuit has a magnetic field associated with it. Ampère showed that a current-circuit C lying in a plane, and enclosing a small area A, has the same magnetic field, at points whose distance from C is large compared with the linear dimensions of the circuit, as a certain elementary magnet or dipole at C. The axis of the equivalent dipole is normal to the plane of C, and its moment is proportional to the area A and the magnitude i of the current. The constant of proportionality depends on the unit in which i is measured; the unit chosen so as to make the constant of proportionality equal to unity is called the *electromagnetic* unit of current. The positive direction of the current is clockwise when viewed along the positive direction of the equivalent dipole. An electromagnetic unit of current is equal to 10 amperes.

Any circuit C, plane or otherwise, and of any size, carrying a current i, can be replaced by a network of currents i, lying on a surface S, of any form, bounded by the circuit, provided that the current i circulates round every element of the network in the same sense. Every internal line of the network then carries the current i in two opposite directions, so that the resultant current is nil, except round the boundary circuit C. Each elementary current of the network, enclosing an area A small enough to be considered plane, is equivalent to a dipole situated at its centre, normal to its plane and of moment iA. By superposing the known fields of all the equivalent dipoles, the magnetic field of the current i round C is obtained; the result is valid at all points not lying in or very near to the surface S; since this surface can be chosen arbitrarily, the magnetic force at any point P whatever can be found by taking S so as not to pass through P.

If all the dipoles are taken to have the same length d, their pole-strengths must be $\pm iA/d$. All the positive poles may be considered as lying on a surface S_+ parallel to S, at a distance $\tfrac{1}{2}d$ on the positive side of S, and all the negative poles on a similar surface S_- at the same distance on the negative side. Since the pole-strength is (i/d) times the area A, when the subdivisions of S are very numerous, and the areas A very small, the distributions of poles over S_+ and S_- tend to

equivalence with two *surface* distributions of pole-strength over these surfaces, with the constant surface-density $+i/d$ over S_+, and $-i/d$ over S_-. Such a combination of equal and opposite uniform surface distributions of magnetic matter over two closely adjacent parallel surfaces is called a *magnetic shell*.

The problem of finding the magnetic field of a current-circuit of any form is thus reduced to the purely mathematical one of integrating the fields of such a distribution of elementary magnets. By methods which are described in treatises on electromagnetism, it is easy to prove the following two theorems due to Ampère.

(i) The magnetic potential V at a point P, due to a circuit C carrying a current i, is equal to i times the solid angle Ω subtended by C at P; Ω is the area cut out, on a sphere of unit radius and centre P, by the conical surface containing C, with vertex P.

(ii) If a closed curve s be drawn in any magnetic field, the line-integral of the magnetic intensity F round s is equal to $4\pi i$, where i is the total current threading the curve, whether in one or more circuits; that is,

$$\oint F \cdot ds = 4\pi i, \tag{24}$$

where, if X, Y, Z are the components of F, and dx, dy, dz those of ds,

$$\oint F \cdot ds \equiv \oint (X\, dx + Y\, dy + Z\, dz).$$

By Stokes's theorem,

$$\oint F \cdot ds = \int \operatorname{curl} F \cdot dS,$$

where the integral on the right is taken over any surface S bounded by the curve s, and where curl F is a vector with the x-, y-, z-components

$$\frac{\partial Z}{\partial y} - \frac{\partial Y}{\partial z}, \qquad \frac{\partial X}{\partial z} - \frac{\partial Z}{\partial x}, \qquad \frac{\partial Y}{\partial x} - \frac{\partial X}{\partial y}.$$

If the electric currents in the field are regarded not as flowing along mathematical lines without lateral extension but as having a finite current-intensity j (i.e. current per unit area of cross-section), then

$$i = \int j \cdot dS,$$

so that (24) may be written

$$\int (\operatorname{curl} F - 4\pi j) \cdot dS = 0.$$

Since this is true for every curve s or surface S bounded by s, it follows that

$$\operatorname{curl} F = 4\pi j. \tag{24 a}$$

This theorem is true whether the field is wholly due to (closed) electric currents, or partly or wholly to magnetic distributions.

The first of these theorems enables the magnetic intensity at any point on the axis of a circular circuit carrying a current i to be calculated simply. Let a be the radius of the current-circuit, and O its centre; let P be any point on the axis, at distance h from O (Fig. 6).
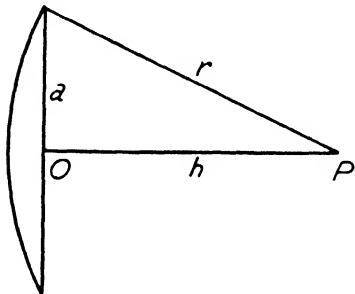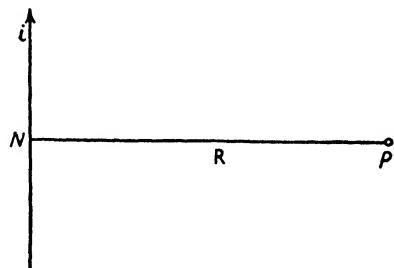
FIG. 6   FIG. 7

Imagine a sphere drawn through the circuit, with centre P, and therefore with radius r given by

$$r^2 = a^2 + h^2.$$

The area of the 'cap' of this sphere bounded by the circuit is $2\pi r(r-h)$, and the solid angle Ω subtended by the circuit at P is this area divided by r^2, i.e. $2\pi(1-h/r)$. Hence

$$V = 2\pi i(1-h/r). \tag{25}$$

The magnetic intensity at P is, by symmetry, along OP, and it is of amount $-dV/dh$, or

$$2\pi i \frac{d}{dh}\left(\frac{h}{r}\right) = 2\pi i\left(\frac{1}{r} - \frac{h}{r^2}\frac{dr}{dh}\right) = \frac{2\pi i}{r}\left(1 - \frac{h^2}{r^2}\right) = \frac{2\pi i a^2}{r^3}. \tag{26}$$

At O this reduces to $2\pi i/a$.

At all points in the plane of the circular current, within the ring, Ω has the same value 2π: and therefore V does not vary in this plane, and F at such points has no component in the plane, that is, the lines of magnetic force cross this plane at right angles; this is equally true, for similar reasons, outside the ring, where $\Omega = 0$.

In the case of a current i flowing along an infinite straight line, the lines of magnetic force must, by symmetry, be circles centred on the line, and lying in planes normal to the line. By applying the second of Ampère's theorems to a curve coinciding with any such line of force, of radius R (Fig. 7), we have

$$2\pi R F = 4\pi i,$$

since any such circle is threaded solely by the current i. Hence

$$F = 2i/R. \qquad (27)$$

Looking along the direction of i, the positive direction round the lines of force is clockwise.

At points very close to a curved current i, such that the distance R is small compared with the radius of curvature, the magnetic intensity
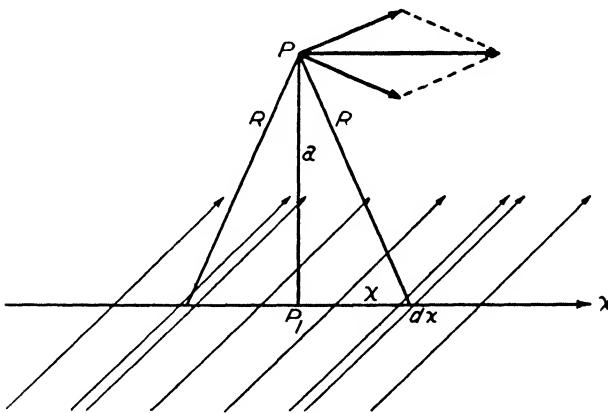
FIG. 8

is nearly the same as if the current were straight, that is, the lines of force are very nearly circles having the current as axis, and $F = 2i/R$. This is true, for example, at points very near the current-ring of Fig. 6, but as the distance from the current increases, the formula ceases to be valid. This is evident from the value of F at O, which is $2\pi i/a$.

A wire wound in a uniform spiral round a circular cylinder is called a *solenoidal coil* or *solenoid*. If a current i is flowing in an infinitely long solenoid with n turns per cm. length (measured along the axis of the cylinder), Ampère's second law gives $H = 4\pi ni$ for the total magnetic force inside the solenoid; this is independent of the radius of the cylinder. If the current-intensity is A amperes, we have to replace i by $A/10$, so that

$$H = 0.4\pi nA. \qquad (28)$$

The product nA is called 'number of ampere-turns per cm.', and the field intensity 1 gauss is produced in a long solenoid with $5/2\pi$ ampere-turns per cm.

A distribution of current over a surface, flowing along the surface (regarded as a thin shell), is called a *current-sheet*. The current-density i at any point P is defined as the current flowing across a narrow line

of length ds, drawn through P in the surface, normal to the direction of current-flow at P, divided by ds.

An important special case is that of a plane uniform current-sheet, in which i has everywhere the same direction and magnitude. The magnetic field of such a current-sheet can be calculated as follows, for any point P *outside* the sheet. Draw (Fig. 8) a perpendicular (length a) from P to the plane, and take the foot of this perpendicular, P_1, as the origin of x-coordinates in the plane, increasing in the direction at right angles to the current, from left to right as seen looking along the current. Imagine the current to be divided up into a number of small portions, flowing along strips of width dx. Each strip can be treated as an infinite linear current of strength $i\,dx$. The contributions made by two such strips, at distances $\pm x$ from P_1, to the magnetic intensity at P_1, are each of magnitude $2i\,dx/R$ or $2i\,dx/\sqrt{(a^2+x^2)}$. The components normal to the plane are equal and opposite, and therefore cancel; the components in the x direction are each equal, and together contribute

$$\frac{4i\,dx}{\sqrt{(a^2+x^2)}}\cdot\frac{a}{\sqrt{(a^2+x^2)}},$$

or $4ia\,dx/(a^2+x^2)$. Hence the total intensity H at P is parallel to the plane, in the x direction, normal to the current, and its amount is given by

$$H = \int_0^\infty \frac{4ia\,dx}{a^2+x^2} = 2\pi i = 0 \cdot 2\pi A, \tag{29}$$

if the current-density is A amperes per cm.

1.14. Intensity of magnetization. Any body which exerts a magnetic field (in addition to the field of any electric currents flowing within* it) is said to be magnetic or magnetized. Imagine that it is subdivided into small volume elements δv_1, δv_2,... situated at points P_1, P_2,...; the whole field of the body is the sum of the fields of these small parts. The field of each element, say δv_1, at points whose distance from P_1 is large compared with the size of the element, is the same as that of an elementary dipole of small magnetic moment M_1 (say) situated at P_1, with its axis in some definite direction. If the subdivision is varied, the elements always remaining small, though containing a large number of molecules, the moment of the dipole equivalent to the element at P_1 will be proportional to δv_1, and its direction will be practically independent of δv_1. Thus $M_1 = I\delta v_1$, where I (a vector, having direction as well as magnitude) is a characteristic of the body in the neighbourhood of P_1, though not necessarily the same at different points. I is called

the (vector) *intensity of magnetization* of the body; it is a function of position in the body. The magnitude of I will be denoted by I; this is the same symbol as is used for the magnetic dip or inclination, but the context will make it clear in which sense the symbol is used.

A uniformly magnetized body is one for which I is the same throughout, in direction and magnitude.

Since the unit for magnetic moment is gauss cm.³ (§ 9), the unit for intensity of magnetization is the gauss, the same as for magnetic force.

The potential V of a magnetized body at any point, inside or outside, is simply the sum of the potentials due to elementary dipoles $I\,\delta v$ distributed throughout the body, and the magnetic force is everywhere

$$F = -\operatorname{grad} V. \tag{30}$$

1.15. The field of a uniformly magnetized sphere. As a consequence of the definition of I, it follows that the potential of a *uniformly* magnetized body is the same as the sum of the potentials for two uniform distributions of magnetic poles throughout the volume, of (large) strengths m, $-m$ per unit volume, the positive distribution being displaced relative to the negative, in the direction of I, by a small distance d, such that $md = I$.

The potential of a uniform *spherical* distribution of gravitational, magnetic, or other matter which exerts a force varying as the inverse square of the distance was shown by Newton to be the same, at outside points, as if the matter were concentrated at the centre of the sphere. For a sphere of magnetic density m and radius a, the equivalent magnetic pole at its centre is therefore of strength $\frac{4}{3}\pi a^3 m$, and the potential at an outside point distant r ($> a$) from O is $\frac{4}{3}\pi a^3 m/r$; the corresponding radial force is $\frac{4}{3}\pi a^3 m/r^2$. It can be shown that at inside points r ($< a$) the part of the sphere beyond the radius r exerts no force, and the force is therefore the same as that produced by the inner sphere, of radius r, at its surface, namely, $\frac{4}{3}\pi r^3 m/r^2$ or $\frac{4}{3}\pi m r$; the corresponding potential (continuous with that at outside points on crossing the surface $r = a$) is $\frac{2}{3}\pi m(3a^2 - r^2)$.

A sphere of radius a and centre O, uniformly magnetized to intensity I in the direction Oz, is equivalent to two such uniform distributions of magnetic matter, of densities $\pm m$, with slightly different centres O_-, O_+, at the distance $\frac{1}{2}d$ on either side of O along the line Oz, and such that $md = I$. At external points the field of each such distribution is the same as that of a single pole of strength $\frac{4}{3}\pi a^3 m$ at its centre;

PLATE 1
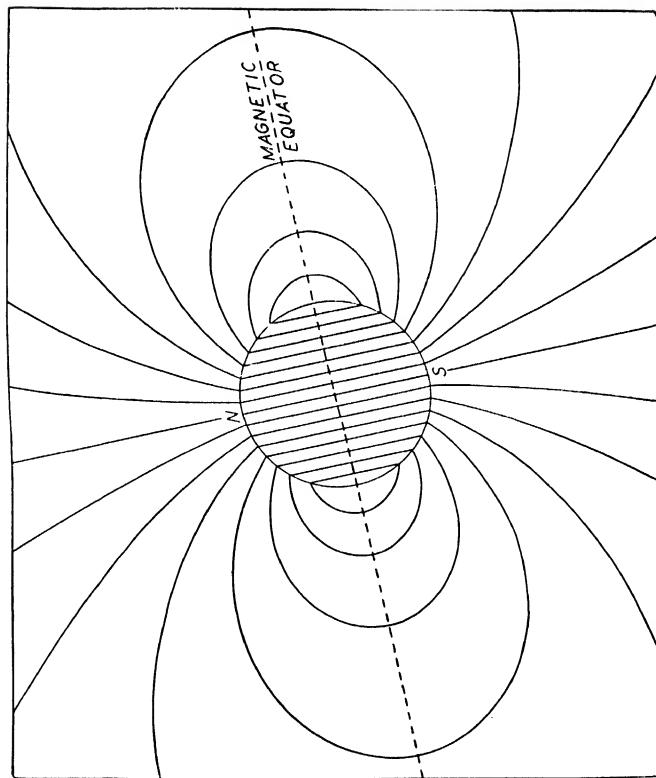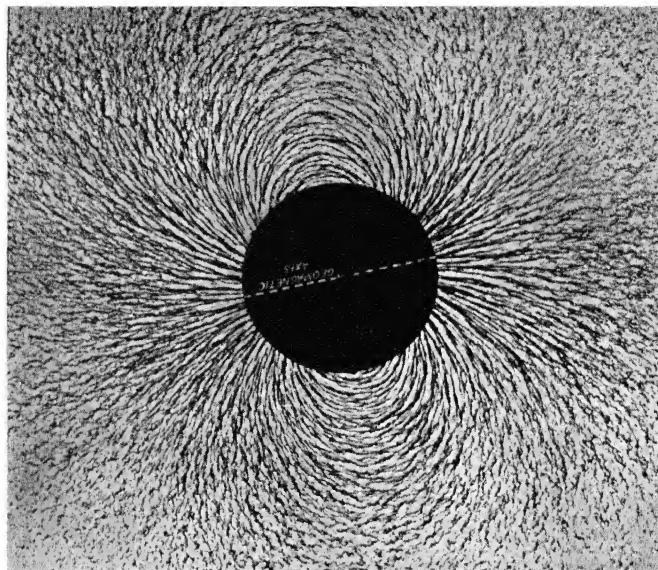
The lines of force for a uniformly magnetized sphere: computed (*left*), and as shown by iron filings (*right*)

hence the two together are equivalent to a dipole of magnetic moment $M = \frac{4}{3}\pi a^3 md$ or $\frac{4}{3}\pi a^3 I$. Consequently, at an external point r, θ (Fig. 9) the potential of the uniformly magnetized sphere is, by (15), $M \cos\theta/r^2$, and the lines of force are given by (17).

At an internal point r, θ the potential is

$$\frac{2}{3}\pi m\{(3a^2 - r_+^2) - (3a^2 - r_-^2)\} = \frac{2}{3}\pi m(r_-^2 - r_+^2)$$
$$= \frac{2}{3}\pi m\{(r + \tfrac{1}{2}d\cos\theta)^2 - (r - \tfrac{1}{2}d\cos\theta)^2\}$$
$$= \frac{4}{3}\pi mrd\cos\theta = \frac{4}{3}\pi Ir\cos\theta = \frac{4}{3}\pi Iz.$$

The corresponding force is $-\mathrm{grad}\,V$, and as V varies only with respect to z ($= r\cos\theta$), the force is $-\frac{4}{3}\pi I$ parallel to Oz. Thus the internal lines of force are as shown in Plate 1.

In the case of the earth, it is usual to reckon the polar distance θ from the north pole N. But the earth appears magnetized in the approximate direction from geographic north to south, because the north-seeking needle is attracted to the northern regions, which must therefore themselves have 'south-seeking' magnetic polarity. It is therefore

FIG. 9

convenient to give the formula for a sphere magnetized *opposite* to the direction ON, or $\theta = 0$. Its potential at a point P at distance $r > a$ from the centre O is clearly $M\cos(\pi - \theta)/r^2$ or $-M\cos\theta/r^2$, where

$$M = \tfrac{4}{3}\pi a^3 I. \tag{31}$$

The radial and tangential components of force, R, S, at P are the reverse of those in (16), namely,

$$R = -\tfrac{8}{3}\pi I\left(\frac{a}{r}\right)^3 \cos\theta, \qquad S = -\tfrac{4}{3}\pi I\left(\frac{a}{r}\right)^3 \sin\theta. \tag{31a}$$

Here, as in (16), R is reckoned positive outwards, and S is positive in the direction of increasing θ, that is, towards geographic south. Near the geographic north pole N (where $\theta = 0$), R is downward; S is everywhere northward. Let the northward horizontal and the downward vertical force components at the surface $r = a$ be denoted by H and Z. Then

$$H = \tfrac{4}{3}\pi I\sin\theta, \qquad Z = \tfrac{8}{3}\pi I\cos\theta.$$

At points on the circle $\theta = 90°$, which may be called the *magnetic equator* (§ 1) of the sphere, H will be denoted by H_0. Then

$$H_0 = \tfrac{4}{3}\pi I, \tag{32}$$
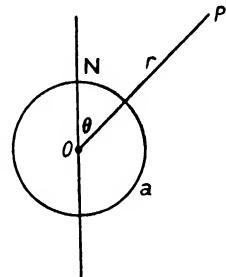
and at other points, in *co-latitude* θ or *latitude* $(90°-\theta)$,

$$H = H_0 \sin\theta, \qquad Z = 2H_0 \cos\theta. \tag{33}$$

Thus at the north pole $Z = 2H_0$, and at the south pole $Z = -2H_0$. The resultant force F in co-latitude θ is given by

$$F = \sqrt{(H^2+Z^2)} = H_0\sqrt{(\sin^2\theta+4\cos^2\theta)} = H_0\sqrt{(1+3\cos^2\theta)}, \tag{34}$$

and the angle ϕ of magnetic dip below the horizontal (this is $90°-\alpha$, where α is the angle defined in (16 a)) is given by

$$\tan\phi = Z/H = 2\cot\theta. \tag{35}$$

The magnetic force inside the sphere is everywhere of intensity $-H_0$, parallel to the *magnetic axis* NO.

The moment M of the equivalent dipole is called the magnetic moment of the sphere; it is expressible as follows in terms of the surface force H_0 at the equator, from (31) and (32),

$$M = H_0 a^3. \tag{36}$$

1.16. The magnetic induction (B) inside a magnetized body. The magnetic force $H = -\operatorname{grad} V$ at any point P inside a magnetized body is that which would be measured if suitably small measuring instruments were placed at P inside a small cylindrical cavity made in the body, centred at P and with its axis along the direction of magnetization (I) at P, provided that the cylinder is very long compared with its breadth (Fig. 10 a). If, however, the cylinder be very short compared with its breadth (Fig. 10 b), the measured force would be different; the force so measured is called the *magnetic induction*, and is denoted by B, with rectangular components B_x, B_y, B_z. It may be shown that B, H, and I are connected by the vector equation

$$B = H+4\pi I. \tag{37}$$

In free space, therefore, $B = H$. (The symbols B and H for magnetic induction and force are widely used in physics, but no confusion need arise from the fact that, in geomagnetism, H denotes the *horizontal* component of magnetic force.)

If we draw any small closed curve C, and imagine that a point P describes this curve from some starting-point upon it, back to this point, the line of magnetic induction through P will generate a tubular surface, which is called a *tube of magnetic induction*. These tubes have the property that at each point along them their normal cross-section

dS is inversely proportional to the induction B, so that $B\,dS$ is a constant for each tube. This property can be mathematically expressed by the equation

$$\operatorname{div}\boldsymbol{B} = 0 \quad \text{or} \quad \frac{\partial B_x}{\partial x} + \frac{\partial B_y}{\partial y} + \frac{\partial B_z}{\partial z} = 0. \tag{38}$$
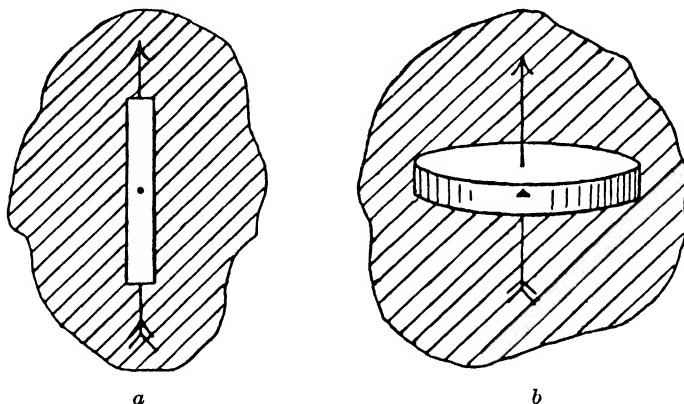
FIG. 10. Magnetic force and magnetic induction inside a
magnetized body

1.17. Permanent and induced magnetization. The magnetization of a body is said to be *permanent* if it is independent of the presence or absence of extraneous magnetic bodies or electric currents and their fields, i.e. if $\boldsymbol{I}$ is the same, at each point P of the body, whether the sole field is that due to the body itself, or whether other fields are superposed. In many substances, however, part of $\boldsymbol{I}$ depends on the value of the magnetic force H at the point, and vanishes when this is made zero; such magnetization is called *induced* magnetization. In isotropic bodies possessing only induced magnetization,

$$\boldsymbol{I} = \kappa\boldsymbol{H}, \tag{39}$$

i.e. the magnetization is in the same direction as $\boldsymbol{H}$; κ, called the *susceptibility*, may be constant or may depend on the magnitude of H.

Combining the equations (37) and (39) for $\boldsymbol{B}$ and $\boldsymbol{I}$, we have

$$\boldsymbol{B} = (1+4\pi\kappa)\boldsymbol{H} = \mu\boldsymbol{H}, \tag{40}$$

where

$$\mu = 1+4\pi\kappa; \tag{41}$$

μ is called the *permeability* of the substance.

By international agreement between physicists, the *gauss* is now adopted as the unit of magnetic induction, and the unit of magnetic force is now called the *oersted*. Since, in air, at sea-level, $B = 1\cdot0000003H$,

we feel at liberty to follow the general usage of geophysicists, who apply the historical unit gauss to the results of geomagnetic measurements, made in air, without reduction [10].

In general a magnetizable body, even if isotropic and homogeneous, does not become uniformly magnetized when brought into a uniform magnetic field, because the induced magnetization at any point depends
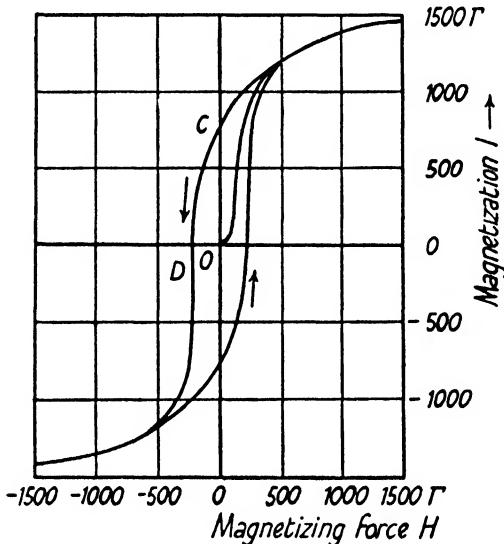
FIG. 11. Hysteresis diagram, showing the magnetization *I* as a function of the magnetizing field strength *H*, for 35 per cent. cobalt permanent magnet steel (after A. M. Armour [9])

on the resultant intensity of the original field and of the secondary field due to the body itself, and the combined field is not uniform. The secondary field, within the body, in general opposes the original field, though not in exactly the same direction. Only for ellipsoids and spheres is the internal field and consequently the induced magnetization uniform; the internal field is of smaller intensity than the primary field. The proportionate internal reduction of the intensity of the primary field may be called the *coefficient of demagnetization*; it varies with the form and permeability of the body.

For most substances the susceptibility κ is of the order $\pm 10^{-6}$; such substances are called *diamagnetic* if κ is negative, or *paramagnetic* if κ is positive.

Certain substances are characterized by high values of κ, and these are called *ferromagnetic*; for such substances, as H increases, κ (or μ)

rises to a maximum value κ_{max} (or μ_{max}), and then decreases in such a way that I attains an upper limit I_{max}.

Ferromagnetic bodies also show the phenomenon of *hysteresis*, that is, I depends not only on the value of H existing at the time, but also on the previous changes of H; for example, if H is reduced to zero, the magnetization does not diminish to zero, but to a value which depends on the earlier strength of the field. Two important quantities associated with this property are the *remanent magnetization* R_0 left after the disappearance of a very strong field, and the *coercive force* H_0 which must be applied in the sense opposite to the primary field in order to neutralize R_0. Fig. 11 shows the magnetization I as a function of the applied magnetizing force H; with fresh material that has not before been magnetized, as H increases from zero the curve begins at the origin O and rises to a maximum, but does not return to O with decreasing H. The ordinate OC represents R_0, and OD gives H_0.

Material at its maximum magnetization is called magnetically *saturated*.

The following approximate figures, after Gumlich [6], will illustrate the wide range in the magnetic properties of commercial iron. (The 'steel' contains 1 per cent. carbon, 0·1 per cent. silicas, 0·4 per cent. manganese, less than 0·1 per cent. phosphorus and sulphur.)

TABLE 1

	Electrolytic iron	Steel	
		Untempered	Tempered
Susceptibility for $H = 1\,\Gamma$ . . .	800	7	3·4
Maximum susceptibility κ_{max} . .	1,140	29	8·6
Maximum magnetization I_{max} in strong fields ($> 1,000\,\Gamma$)	1,720	1,570	1,420 Γ
Remanent magnetization R_0 . .	870	1,040	600 Γ
R_0 as a percentage of I_{max} . . .	51	66	42
Coercive force H_0	0·3	17	52 Γ
Ratio H_0/κ_{max}	3×10^{-4}	0·6	6·0 Γ

Thus electrolytic iron easily gains a high magnetic moment (or is magnetically 'soft'), and can retain a considerable portion of it, but the remanent magnetization is easily destroyed, even by fields not stronger than that of the earth. The material permalloy, consisting of 78 per cent. nickel, 22 per cent. iron, has for small fields of 0·2 Γ a susceptibility five times higher than that of electrolytic iron, with equally small values of R_0 and H_0. Steel, especially if it is *tempered* (that is, heated and cooled successively in an appropriate way), is hard

to magnetize, but it keeps its remanent moment even under the action of strong demagnetizing fields.

For geomagnetic measurements, permanent magnets were formerly made of steel containing a small amount (about 6 per cent.) of tungsten, because of the high values of H_0 (= 70 Γ) and R_0 (= 800 Γ) for this material. It is now outclassed by steel containing 35 per cent. of cobalt; this alloy, when tempered, has about the same remanence as tungsten steel, but its coercive force is three times as great as that for tungsten steel. Johnson [2.92 c] recommends Alnico.

The figures in Table 1 are obtained from experiments which realize practically the ideal case of bodies in the form of closed rings; since no poles arise in rings, complicated demagnetizing effects are avoided. In the bar-magnets used in observatories, however, the form and dimensions are such that strong poles arise and demagnetization is considerable. So the *apparent* remanent magnetization, measured, for example, for bars of 20 cm. length and 1 cm.² cross-section, is only a fraction of that measured for rings. With a sample of untempered steel ($H_0 = 3 \cdot 4\ \Gamma$), R_0 fell from 620 Γ in a ring to 32 Γ in bar-form; with tungsten steel ($H_0 = 72\ \Gamma$), R_0 fell only from 850 to 550 Γ [6]. The high coercive force of cobalt steel ($H_0 = 230\ \Gamma$) withstands the demagnetizing forces even better, so that permanent magnets of a given moment, when made of cobalt steel, may be only one-third as long as ordinary steel magnets of the same cross-section. This improves the accuracy of measurements, because shorter magnets better represent the ideal case of elementary magnets, with a deviation constant $k = 1$ (§ 12).

If the temperature rises above a certain degree, all substances lose their ferromagnetism. Metallurgical investigations show that there are several points at which iron changes its internal structure and, consequently, its magnetic properties. But one point is specially important, because the drop in magnetization is specially marked, namely, for iron 780° C., magnetite 580° C., nickel 350° C.; it is called the *magnetic inversion temperature*, or *Curie point*. Above these respective temperatures the substances practically cease to be ferromagnetic.

Metallurgical research [5] has developed many alloys such as μ-metal, permalloy, perminvar, etc., with special magnetic properties which may become useful for geomagnetic measurements.

1.18. Electromagnetic induction. In a varying magnetic field electromotive forces arise, and electric currents are induced in any electrically conducting bodies which may be present. The direction

and intensity of these currents depend on the rate of variation of the magnetic field and the distribution of electrical conductivity. The calculation of the currents is based on (i) Ohm's law, that the current round any circuit is equal to the total electromotive force round the circuit, divided by the total electric resistance, and (ii) Faraday's law, that the total electromotive force E is equal to the rate of change with time t of the total magnetic induction through any surface S bounded by the circuit, i.e.

$$E = -\frac{d}{dt}\int \boldsymbol{B}\cdot \boldsymbol{dS} = -\frac{d}{dt}\int B\cos\theta\, dS, \qquad (42)$$

where the surface-integral is taken over the surface S, and $\boldsymbol{dS}$ denotes $\boldsymbol{\nu}\, dS$, where $\boldsymbol{\nu}$ is the unit vector normal to the surface-element, of area dS; θ denotes the angle between the normal $\boldsymbol{\nu}$ and the direction of $\boldsymbol{B}$ at that point. The electromagnetic unit (e.m.u.) for E equals 10^{-8} volt.

If a rigid circuit of any form is rotated in a uniform magnetic field about an axis parallel to the field, the total magnetic induction through it remains constant, and therefore the electromotive force E round the circuit is zero, and no current is induced in it; but if the axis is not parallel to the field, the magnetic induction varies periodically, and an alternating electromotive force and current are induced in the circuit.

The total electromotive force E round a fixed *horizontal* plane circuit, enclosing an area S, due to variations in a uniform magnetic field, depends only on dZ/dt, the rate of variation of the vertical intensity Z. Clearly

$$E = -S\frac{dZ}{dt} \text{ e.m.u.} \qquad (43)$$

For example, if $dZ/dt = 10^{-5}$ gauss/sec. $= 1\gamma$/sec., and S is 1 sq. km. or 10^{10} cm.^{2}, $E = 10^5$ e.m.u. $= 10^{-3}$ volt; if the resistance of the circuit is 100 ohms, a steady electromotive force of this amount would impel a current of 10^{-5} ampere round the circuit.

Faraday's electromotive force can be regarded as the line-integral of an electric field induced by the relative motion of matter and the magnetic field $\boldsymbol{F}$. If the latter, or the bodies producing it, are regarded as stationary, and the matter is in motion with velocity $\boldsymbol{v}$, the induced electric field is $\boldsymbol{v}\wedge\boldsymbol{F}$. If the rectangular components of $\boldsymbol{v}$ are u, v, w, the corresponding components of the induced electric field are

$$vZ-wY, \qquad wX-uZ, \qquad uY-vX. \qquad (44)$$

This field can be called a *dynamo* field, since it is a field such as is operative in the armature of a dynamo.

If v is measured in cm./sec., and F in gauss, the induced electric intensity $v \wedge F$ is obtained in electromagnetic units; 1 c.m.u. of electric intensity is 10^{-8} volt/cm. or 10^{-3} volt/km.

An electric particle of mass m and charge e, moving with velocity v in a magnetic field F, is subject to a deflecting force $ev \wedge F$. In a rectangular, right-handed coordinate system, chosen so that F is along z, and v is along x, the particle is accelerated along the negative y-axis, and

$$m\ddot{y} = -euF. \tag{45}$$

In this equation e is supposed expressed in electromagnetic units; for example, if the charge is that of an electron, its numerical value in electrostatic units is

$$4\cdot77 \times 10^{-10} \text{ e.s.u.}; \tag{46}$$

this must be divided by c, the velocity of light, given by

$$c = 3\cdot00 \times 10^{10} \text{ cm./sec.},$$

in order to obtain the electromagnetic charge corresponding to (46), namely

$$1\cdot59 \times 10^{-20} \text{ e.m.u.} \tag{47}$$

We may also note here the values of m for an electron (m_e) and for a hydrogen atom (m_H), which are

$$m_e = 9\cdot01 \times 10^{-28} \text{ gm.},$$

$$m_H = 1\cdot662 \times 10^{-24} \text{ gm.}$$

MAGNETIC OBSERVATIONS

A. PRINCIPLES AND ORGANIZATION OF MEASUREMENTS

2.1. Accurate measurement of a weak magnetic field. The earth's magnetic field differs from those met with in engineering, or in physical research, in two ways. (i) Whereas the latter range in intensity up to 50,000 Γ, or even (in very small spaces and for very short times) 1,000,000 Γ, the intensity of the earth's field is only 0·3 to 0·7 Γ at the surface, except in a few areas where, owing to local magnetic deposits, the intensity has abnormal values up to about 4 Γ (cf. 4.11). (ii) Further, even to the high order of accuracy aimed at in geomagnetic measurements, the earth's field is in general uniform throughout a space of some cubic metres round any point. For this reason the resultant force (tending to move the mass-centre of any ordinary magnet) is extremely small (cf. 1.10), less than one-millionth of its weight; consequently the intensity of the field is usually measured by its rotatory effect, that is, by the *couple* it exerts on a needle. The attempt is made to measure the time-changes of the earth's magnetic intensity F at any point with an accuracy of 1 γ, or about one part in 50,000; this proportion is equivalent to 0·002 mm. in a length of 10 cm., or to 2 seconds in a day. It is difficult to attain this high accuracy in the measurement of the whole intensity, though this standard is aimed at and, at some magnetic observatories, approached.

Obviously it is useless to seek such accuracy unless care is taken that the earth's field at the point where measurements are made is undisturbed by artificial causes, such as underground iron pipes, iron fittings in the observing room, or iron articles in the pockets or clothes of the observer; or by stray electric currents, especially direct currents, in lighting or heating circuits in the building, or from electric traction systems in the vicinity (alternating currents with more than 10 cycles per second are usually innocuous) [83a]. The observing instruments themselves must be constructed of brass or other strictly non-magnetic material, because any trace of magnetic material will be disastrous to accuracy; this necessity cannot be stressed too much (§ 40).

The same principles of measurement are used, but in rather different ways, in observatory work, in general surveys by land and sea, and in local surveys for prospecting (discussed in Chapter IV).

No attempt will be made here to give minute details of magnetic

instruments or methods of observation, for which the publications of observatories or surveys, or of geomagnetic instrument makers, should be consulted [3]. A brief account will be given, however, of the main principles of geomagnetic measurement in the present part A, §§ 1–20, of this chapter; a few general topics will be discussed in part B, §§ 21–40.

2.2. Magnetic observatories. The most accurate and complete information regarding the magnetic force **F** at individual stations is
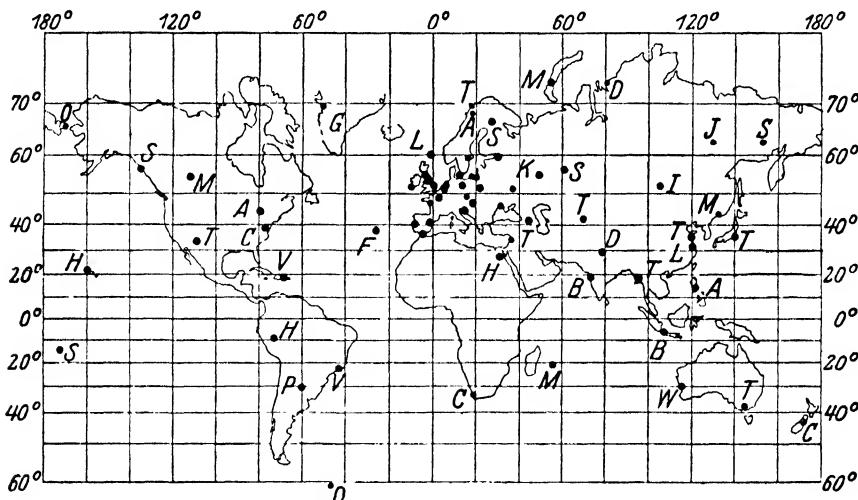
FIG. 1. World distribution of magnetic observatories, 1933

derived from magnetic observatories, of which there are now about seventy-five. Particulars are given in Table A at the end of this book. Their distribution over the earth is indicated in Fig. 1 (for 1933; since then a few more have been established): they are numerous in Europe and also well distributed in America, but the rest of the world is sparsely supplied. The southern hemisphere is shown as having only 11, and the polar regions, beyond 70° latitude, only 2. The difficulty of maintaining observatories in high latitudes is naturally very great, and most of the available information about these regions is derived from observatories established for a period of one or two years only, in association with polar expeditions. In 1882–3 and 1932–3 many nations combined to establish observatories in the Arctic regions for about a year (see § 17), and their simultaneous observations have proved of great value. During the present century British, Australasian, and German expeditions [G 94 ff.] to the Antarctic have included magnetic

observatory work in their programme. The magnetic variations in the polar regions prove to be specially intense and interesting.

Absolute observations. At a magnetic observatory F is determined at regular intervals as accurately as possible. The elements usually measured are H, D, and I, but in high latitudes, where H is small, Z, D, and I may be measured. Such measurements are called absolute observations—a name which dates back to a famous paper by Gauss [1], who was the first to express a non-mechanical quantity, the magnetic force, in the absolute units of length, time, and mass. These absolute observations are made monthly, weekly, or oftener; the practice in this respect varies considerably from one observatory to another.

2.3. Absolute observations of direction: declination and dip. The declination D is measured by observing the azimuth of a needle freely suspended by a practically torsionless thread, attached at such a point, differing from the mass-centre, that the magnet hangs with its magnetic axis horizontal. The direction of geographic north, from which D is measured, must naturally be determined by astronomical or geodetic observations in the first instance, and it is preserved by suitable marks in the vicinity with accurately known azimuths. The declination is the simplest absolute observation, but great skill and care are required in order to obtain the result with the desired accuracy of $0 \cdot 1'$.

In order to approach the ideal condition of suspension by a torsionless fibre, a thin fibre and a strong magnet (of moment M) are used. The upper end of the suspension-fibre can be turned in a torsion-head with a circular scale which indicates its position relative to the suspension-tube. The torsionless state is found approximately by suspending a non-magnetic bar; the torsion-head is then turned till this bar is nearly in the magnetic meridian. This adjustment (position A of the torsion-head, say) will not be perfect, so that the torsion-head will differ from its correct position by an angle θ, say, of a few degrees. When the bar is replaced by the magnet, this residual twist in the fibre will deflect the magnet by a small angle from the meridian; hence the azimuth α of the magnet, relative to the fixed azimuth mark, will differ slightly from that, α_0, for a torsionless fibre. In order to eliminate this residual torsion effect, a second azimuth reading, α', is taken for a weaker magnet (of moment M', between $\frac{1}{3}M$ and $\frac{1}{2}M$) suspended by the same fibre.

The ratio of the two deflexions, $(\alpha - \alpha_0)$ and $(\alpha' - \alpha_0)$, and thereby α_0, is determined experimentally as follows. The restoring couples

exerted by the earth's magnetic field on the two magnets are, for small deflexions (see 1.11), $MH(\alpha-\alpha_0)$ and $M'H(\alpha'-\alpha_0)$; the deflecting couples exerted by the torsion are $\tau[\theta-(\alpha-\alpha_0)]$ and $\tau[\theta-(\alpha'-\alpha_0)]$, where τ denotes the torsion-coefficient of the suspension-fibre; since the deflexion angles of the magnets will be small compared with the angle of twist θ, the couples may both be taken as equal to $\tau\theta$. Since the restoring and deflecting couples are in equilibrium, it follows that also

$$MH(\alpha-\alpha_0) = M'H(\alpha'-\alpha_0) = \tau\theta, \tag{1}$$

$$(\alpha-\alpha_0):(\alpha'-\alpha_0) = M':M. \tag{2}$$

The ratio $M':M$ is determined by taking, for each magnet, an additional pair of azimuth readings of the two magnets, δ and ϵ for M, and δ' and ϵ' for M', when the torsion-head is turned through a large angle L (say $60°$), read on the scale of the torsion-head, first towards the left and afterwards towards the right, from its position A; L is chosen so that $(\delta'-\epsilon')$ is not larger than a few degrees, and the sines can be replaced by the angles. The two equations which result from (1) by replacing α and α' by δ and δ' or by ϵ and ϵ' yield, on subtraction,

$$MH(\delta-\epsilon) = M'H(\delta'-\epsilon'),$$

or $\qquad\qquad M':M = (\delta-\epsilon):(\delta'-\epsilon'). \tag{3}$

With this value, (2) is solved for α_0 and yields

$$\alpha_0 = \alpha-(\alpha'-\alpha)M'/(M-M'). \tag{4}$$

The azimuth readings of the magnet—or, more exactly, of the perpendicular to a *mirror* attached to the magnet (for instance, its polished end-surface)—are taken by means of a magnetic theodolite (described in §4) in which the magnet is suspended as the needle. The optical perpendicular to the mirror may not be correctly alined with the magnetic axis of the magnet. Therefore the magnet—which is usually a hollow cylinder—bears a mark on its circumference near its middle, and is suspended in two positions, designated 'mark up' and (the reversed position) 'mark down' (or, also, 'erect' and 'reversed'); the average of the readings in these two positions is the correct azimuth of the magnetic axis (Fig. 2).

The dip I was formerly, and often still is, measured by a *dip-circle* (Plates 2a, 2b), consisting of a magnetic needle which can swing about a horizontal axis through its mass-centre. By means of an auxiliary horizontal compass-needle, the instrument is rotated so that the axis of the dip-needle is perpendicular to the magnetic meridian, when the needle is free to aline itself along the direction of the force F; it swings

parallel to a vertical divided circle, from which its inclination to the horizontal is read. Just as when measuring the declination, it is necessary to take readings in several reversed positions of the needle, to eliminate small errors; even the magnetization of the dip-needle must be reversed in the course of each set of measurements, in order to eliminate eccentricities between the mass-centre and the axis. But the errors of the dip-circle are greater and more difficult to avoid than the
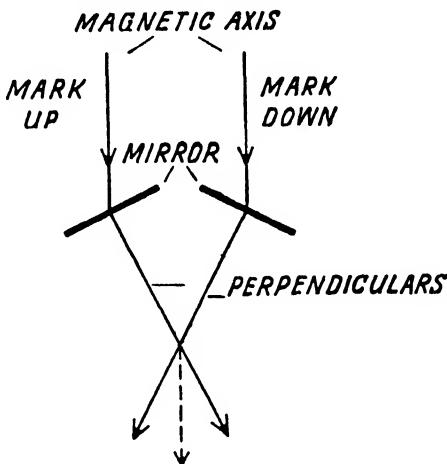
FIG. 2. Schematic plan of a declination magnet
with its attached mirror, suspended in two
reversed positions

errors encountered in measuring the declination, because of the large effect of small departures of the axis from a truly cylindrical form, and the friction of the axis of the needle on the pivots.

Consequently the dip-circle has been gradually superseded by another instrument, the *dip-inductor* or *earth-inductor*, which is capable of measuring I to an accuracy of $0.1'$. It consists of a circular coil of many windings, that can be rapidly rotated about an axis lying along a diameter of the coil. The axis is carried in a frame, and its direction can be varied at will, and measured. If the axis does not lie along the direction of the earth's field, then on rotating the coil an alternating current is induced in it (1.18), which is rectified by a commutator and indicated by a galvanometer. The observation consists in adjusting the frame, while the coil is in rotation, till there is no current; the axis is then in the direction of the field, and its dip below the horizontal is read, by means of microscopes, from the vertical divided circle. Two types are shown in Plate 3, and cross-sections in Fig. 3.
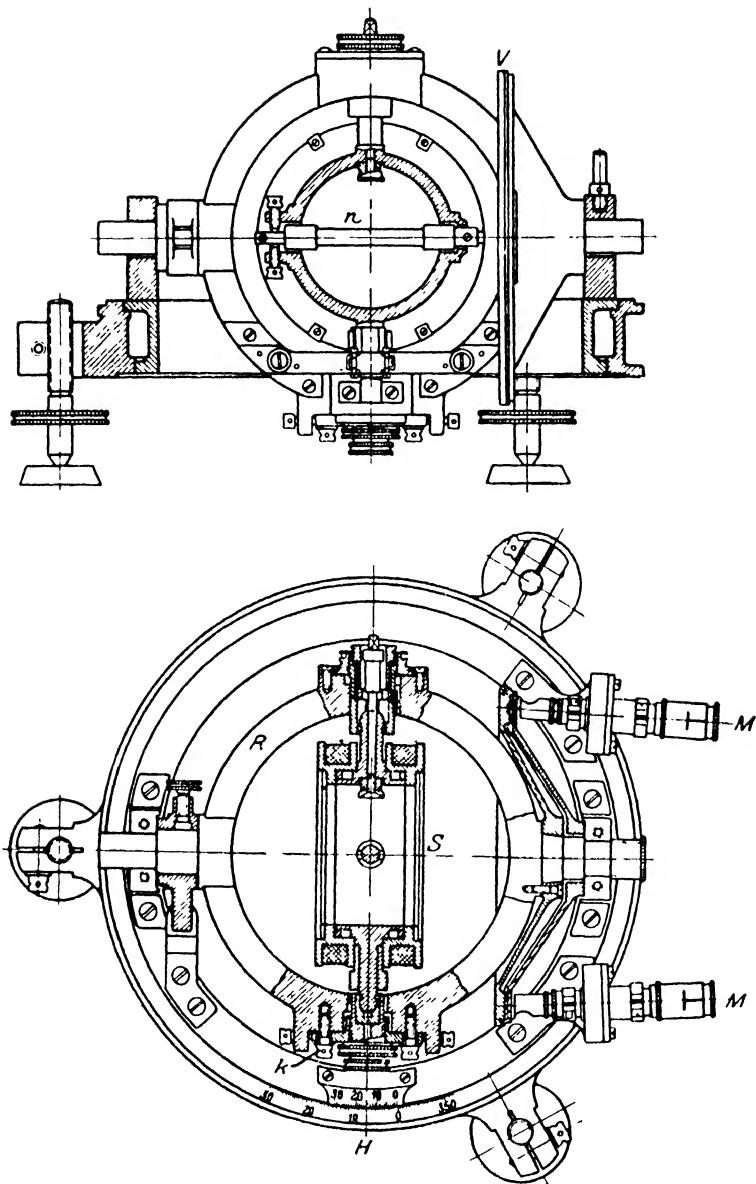
Fig. 3. Schematic sectional elevation and plan of an earth-inductor (Askania)
Coil S, ring R, vertical and horizontal divided circles V and H, microscopes M

The rectification of the current by means of commutators introduces thermo-currents and other parasitic voltages which limit the sensitivity. E. A. Johnson [11] has therefore proposed alternating-current methods of detection, which may be made a thousand times as sensitive as

direct-current methods, and are specially suitable for marine observations.

2.4. The magnetic method of measuring the horizontal intensity H. The horizontal intensity H is still mostly measured by the classical method due to Gauss. This involves two experiments, called the *oscillation experiment* and the *deflexion experiment*. In the former the period of oscillation of a magnet is determined, when suspended horizontally by a fine thread of negligible torsion; about a hundred swings are observed, and the measurement of the time of swing, T, indicates the value of MH/I, where M denotes the moment of the magnet and I its moment of inertia: for, by 1 (21),

$$MH/I = 4\pi^2/T^2. \tag{5}$$

The magnet is not of quite simple form, partly because it may have a mirror or lens attached to it for observation of its swing; hence I cannot be calculated from the mass and size, but must be determined by experiment. The magnet is therefore made so that a cylindrical non-magnetic bar can be attached to it, this bar being of known mass and simple form, so that its moment of inertia I′ can be calculated: for instance, the magnet may be a hollow cylinder into which the bar can be placed. With the bar attached, the whole moment of inertia is I+I′. Observation of T', the time of swing of the magnet plus bar, in the same magnetic field H, thus indicates the value of $MH/(I+I')$. Since MH is the same in the two cases, the ratio (I+I′)/I is known, and from this and the known value of I′ the value of I for the magnet can be inferred. When this has been done, any oscillation experiment with that magnet indicates the value of MH at the time and place of the experiment.

In the deflexion experiment the same magnet is placed horizontally, with its centre O at a distance r from that of a freely suspended horizontal needle, and with its axis directed towards the centre O′ of the needle. The needle is therefore subject both to the earth's horizontal intensity H and to the force $2kM/r^3$ of the magnet M (1.12); its deflexion from the position of rest along the earth's field, when M is brought into position, affords a measure of the ratio $2kM/Hr^3$. In this experiment the movements of the needle are damped by a light vane moving in oil, or electromagnetically by copper dampers.

The deflexion constant k, nearly equal to unity, is supposed to be known, being determined by observing the deflexions of the needle when the magnet M is placed in different positions (§ 30); if it were possible

it would be convenient to make k equal to 1, within the degree of accuracy desired, by making r sufficiently large; but with the intensity of magnetization attainable in ordinary magnets, this would reduce the deflexion at O' too much, so that k must, in fact, be taken into account.

Gauss made the deflexion experiment with the magnet M in a fixed position at right angles to the magnetic meridian; in this case (1 (22)) the deflexion α is given by

$$\tan\alpha = 2kM/Hr^2. \qquad (6)$$

Lamont placed the magnet so that it was perpendicular to the needle after the deflexion; by 1 (23), the deflexion β is then given by

$$\sin\beta = 2kM/Hr^3. \qquad (7)$$

Since k and r are known, when α or β is measured, M/H is determined.

Combining the results of the two experiments, which determine MH and M/H, the separate values of M and H are found. In fact, from (5) and (7),

$$H^2 = C^2/T^2\sin\beta, \qquad (8)$$

where C^2, given by

$$C^2 = 8\pi^2kI/r^3, \qquad (9)$$

is an instrumental constant.

Lamont's method of making the deflexion experiment has superseded that of Gauss, because the results are less influenced by small deviations of the experimental conditions from those intended.

Lamont's method implies the use of a special magnetic theodolite (Fig. 4). The large divided circle B is mounted horizontally on three screws. The circular plate A, supporting all the other parts, may be turned round a vertical axis, and its position can be read off on the divided circle B (on which the divisions are about $\frac{1}{6}°$), by means of verniers, to 1', or preferably by micrometer-microscopes to 2''. Over the centre of A the needle is suspended from an adjustable torsion-head by means of a long thin thread. The needle carries a small plane mirror m, which reflects the image of the threads in the eye-piece of the telescope T, so that if the directly visible threads coincide with the reflected threads, the optical axis of the telescope is perpendicular to the mirror and therefore (with suitable adjustment) parallel to the magnetic axis of the needle. Perpendicular to the telescope, and in the same horizontal plane as the needle, two long horizontal arms are fixed to the box which contains the needle. These arms have grooves for the reception of the deflecting bar-magnet. The distance from the needle is defined by placing the bar in contact with solid stops. In order to eliminate several possible eccentricities, the deflexion β is observed in

four positions of the bar-magnet, with the bar east or west of the needle, and with the north pole of the bar east or west; the mean of the four values of β is taken. The same theodolite is also used for measuring the declination; the needle is then replaced successively by the two magnets mentioned in §3; the magnetic theodolite is also used as an
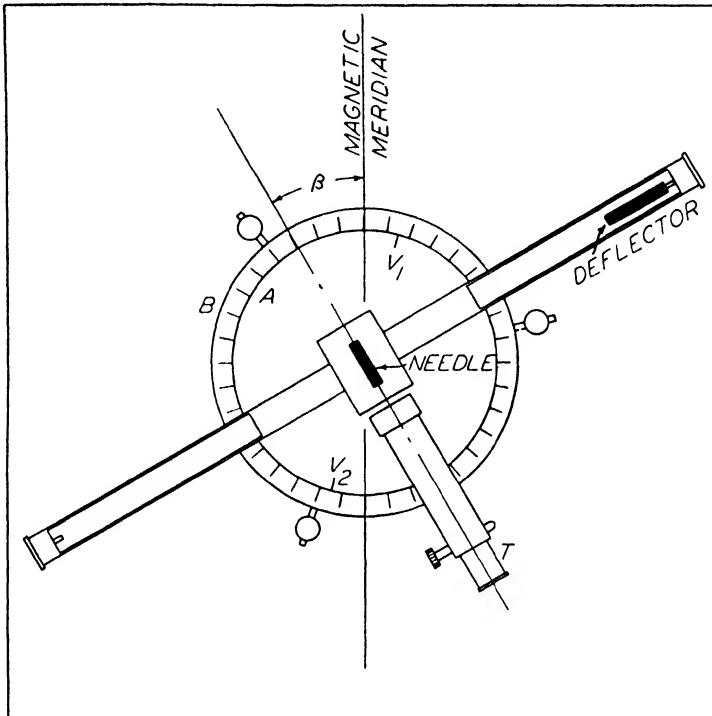
FIG. 4. Schematic plan of a magnetic theodolite

ordinary theodolite for the geodetic or astronomical part of the D measurement.

The instrumental constants, I, k, and r, must be determined with great accuracy. It is also necessary to determine the *temperature coefficient* of the constant C^2, because the complete determination of H takes some time, and the magnet and apparatus may not remain at the same temperature throughout. The moment of an ordinary magnet decreases by about 1/10,000 of its whole amount per 1° C. rise of temperature; consequently, in determining H to 1γ (that is, to about 3 parts in 10^5), it is desirable to know the temperature of the magnet to 0·1° C. Temperature also influences the moment of inertia of the magnet and the length of the arm carrying the magnet in the deflexion experiment.

Another circumstance that has to be taken into account is the effect of the earth's field on the moment of the magnet in different positions. If H' is the component of the earth's field along the axis of the magnet in any position,

$$M = M_0(1+k'H'),\tag{10}$$

where M_0 and k' are constants, M_0 being the value of M when the magnet is transverse to the earth's field. With ordinary magnets the *induction coefficient* k' is about 0·01, so that when the magnet is alined along a field of intensity 0·2 Γ, say, the moment is 0·2 per cent. greater than M_0, or 0·4 per cent. greater than in the reversed position. The change in M from the oscillation experiment, when the magnet is directed along H, to the deflexion experiment, where it is nearly transverse to it, must be allowed for in the determination of H. It is not usually necessary to take account of any transverse magnetization induced in the magnet.

The moment M, as well as the horizontal intensity H, is determined in each complete experiment. The value of M should be very nearly constant, though slightly decreasing with age; any marked change from one experiment to another indicates either some inaccuracy in one of the experiments, or possibly a (non-linear) variation of H during the experiment (unless the magnet has meanwhile been shaken, or heated, or exposed to a strong magnetic field, any of which may permanently change M: all such changes should be carefully avoided).

Since the measurement of the instrumental constant C (involving I, k, and r), and of its temperature coefficient, requires extreme accuracy and a long series of observations, it has been made at comparatively few observatories. Instruments for which the constants have been thus determined are true absolute instruments (Plate 4 a). They may be used as standards by means of which other, *secondary*, sets of apparatus may be calibrated. This is done by observing H with the standard instrument, measuring T and β, and at the same time and place measuring T_1 and β_1 for the secondary instrument. Then, C^2 being the known constant of the standard instrument, and C_1^2 the constant (to be determined) of the secondary apparatus, we have from (8)

$$H^2 = C^2/T^2\sin\beta, \qquad H^2 = C_1^2/T_1^2\sin\beta_1.\tag{11}$$

Hence

$$C_1^2 = (T_1^2\sin\beta_1/T^2\sin\beta)C^2.\tag{12}$$

For rough survey work, H is sometimes measured by means of the deflexion experiment alone; from (7),

$$H = (2kM/r^3)/\sin\beta,$$

PLATE 4

b. A sine-galvanometer. After S. J. Barnett
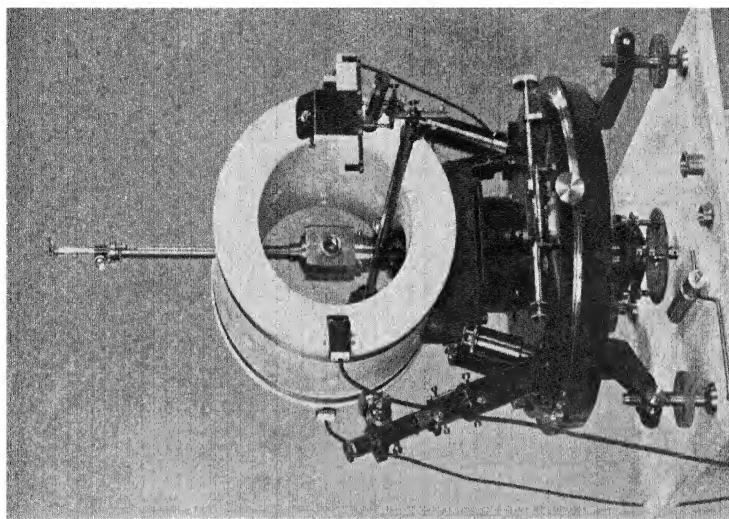
a. A magnetic theodolite (Askania)

1, the annular tripod, 2, the foot-screws, 3, the divided circle, 4, the microscopes, 5, the magnet housing, 6, the deflecting magnet on its supporting arm, 7, the telescope, 8 and 9, the clamp and screw for fine adjustment (for movement around the main vertical axis), 10 and 11, the clamp and screw for fine adjustment of the telescope about the horizontal axis, 12, counter-weights, 13 and 14, the clamping device for the needle, 15, the suspension tube, 16–18, the torsion head with its scale and adjustment device, 19, the level, 20, the box for the magnets, 21, the box for the needle and accessories
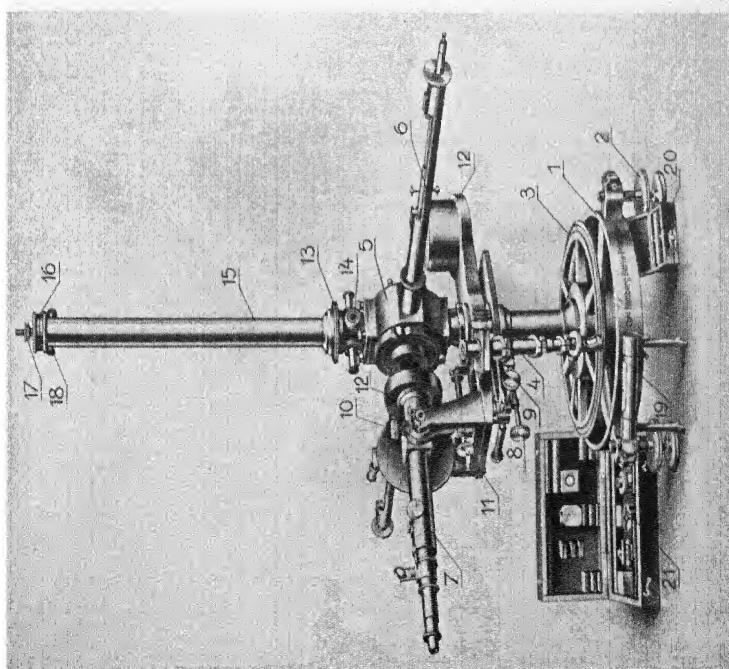
PLATE 5
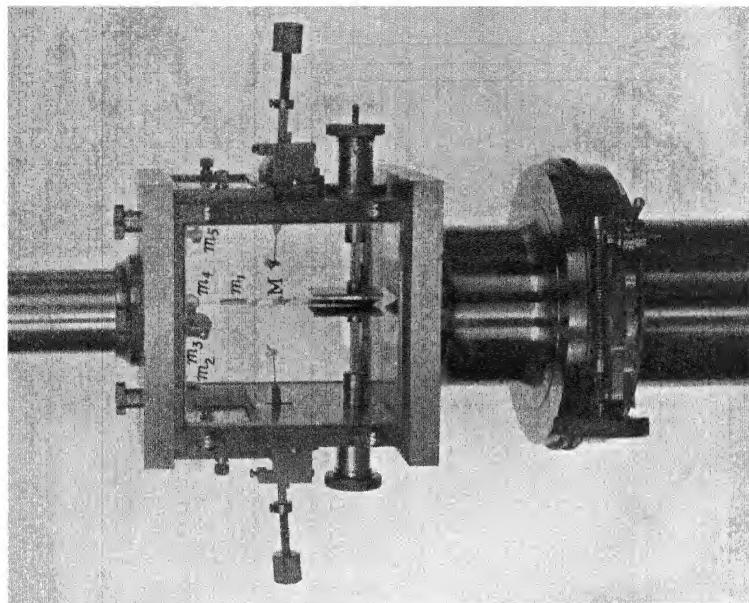
b. View of the magnet housing at the centre of the coil shown on the left

M is the magnet, *m*, the mirror attached to the fibre, m_2 to m_5 are mirrors attached to the housing, to indicate movements of exactly 90°, 180°, etc. A further mirror attached to the marble cylinder by a V-shaped metal bracket is shown on the left, above the magnet housing
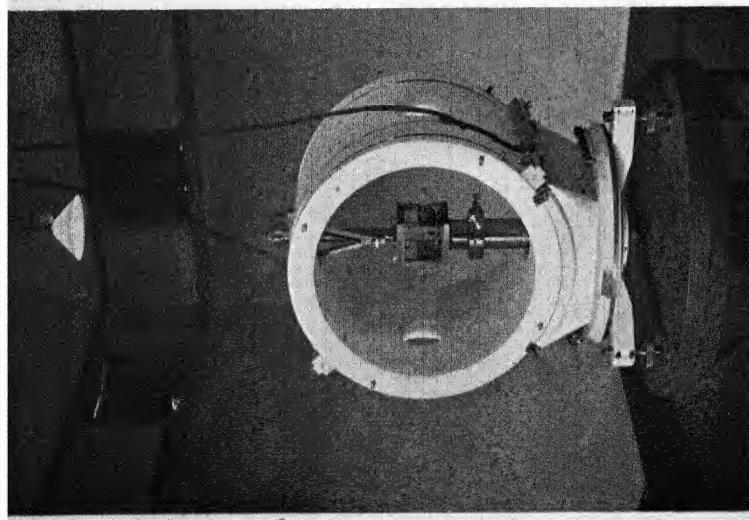
a. General view of the Schuster-Smith coil magnetometer

where $(2kM/r^3)$ is treated as a known constant, determined by complete absolute observations made at suitable intervals (to allow for the slow change of M, or for possible accidental changes in k, M, or r).

Instrumental details regarding magnetic theodolites and the magnetic measurement of H will be given in § 33.

2.5. Electrical methods for measuring magnetic intensities. Owing to the number of difficult measurements needed in determining H very accurately by the magnetic method, there is a growing tendency to base magnetic measurements on the use of electric currents instead of on methods involving permanent magnets. The magnetic field of a current-system is determined by its geometrical arrangement (including the number of windings) and the current-strength i, so that, when the dimensions have been measured with the necessary accuracy, the field corresponding to a given value of i can be inferred. The electrical measurement of H may consist either of a deflexion experiment in which the deflecting field is produced electrically instead of by a permanent magnet, or of the annulment of the earth's field at the centre of the instrument. Usually the current-system consists of two exactly similar coaxial circular coils, each of a few windings, with their planes separated by a distance equal to the radius of the coils; as shown by Helmholtz and Gaugain, this arrangement produces a field which is very nearly uniform over a fairly large space round the point C midway between the centres of the coils (see § 34). The deflected needle is placed at C, and its deflexion when the current i is switched on is measured as in Lamont's method. The current i is measurable by means of a Weston cell to one or two parts in 100,000. A *sine-galvanometer* of this type was designed by S. J. Barnett [23] for the Carnegie Institution of Washington (Plate 4 b). This and other sine-galvanometers are gradually replacing the older magnetometers, although the magnetometer method has, in other respects, the advantage that it is based directly on measurements of length and time, and it should therefore not be superseded entirely.

F. E. Smith [24], developing a plan proposed by A. Schuster [22], designed an extremely accurate *coil-magnetometer*, two copies of which are in use at Abinger (Plate 5 a, 5 b). The Helmholtz coil of the original instrument consists of two short windings, of 12 turns each, of $\frac{5}{6}$ mm. pitch, made of bare copper wire wound in spirals upon a marble cylinder of diameter 60 cm. With its axis horizontal, this system is supported upon a base-plate and can be revolved about a vertical axis. A small needle (1 cm. long, 5 sq. mm. in cross-section) is suspended freely at

the centre of the system; let the intensity at this point, due to the coil, be denoted by H_c.

The current i is adjusted so that H_c is slightly greater than H, and the coil is turned round the vertical axis so that its own axis makes an angle of nearly 180° with the direction of H, this angle being chosen so as to bring the needle at right angles to the magnetic meridian (Fig. 5). Then, if α is the angle which H_c, or the axis of the coil, makes with the reverse direction of H,

$$H = H_c \cos \alpha.$$

If α is kept small (about 2° or 3°), $\cos \alpha$ is very nearly unity, so that small errors in α do not matter. The whole observation may be made in a few minutes.

The two copies of the Schuster-Smith magnetometer, constructed and calibrated quite independently, give mean values for H which differ by only $0 \cdot 6 \gamma$. Thus it is fairly safe to conclude that the values of H measured by these instruments are exact to about 1γ. Later copies of these instruments have been made on a smaller scale, with coil-diameter 50 cm.

FIG. 5. To illustrate the method of observation with the Schuster-Smith coil-magnetometer

In high latitudes, within about 11° distance from the principal poles of magnetic dip (1.1), H is less than $0 \cdot 1 Z$, and F is nearly equal to Z. Therefore in these regions, in order to avoid the large errors (§ 12) which might arise in deducing Z from H and I, it is better to measure Z directly. Two different types of instruments have recently been developed for this purpose.

For the Danish observatory at Godhavn (Greenland), which is a permanent magnetic station near to the magnetic pole, D. la Cour [25] has designed a Z-magnetometer based on the following idea.

An earth-inductor is constructed with two parallel coils, A and B; the coil A is connected with a ballistic galvanometer. Initially the plane of the inductor windings is horizontal; it is then turned through 180° about a horizontal axis. The vertical intensity Z induces in A a current which can be measured by the galvanometer. Initially an auxiliary current i_0, accurately measured, is flowing through the coil B. During the half-turn of the inductor, i_0 is reduced automatically to

zero in suitably graded steps; i_0 is adjusted so that, in the initial horizontal position, it sends twice as many lines of magnetic force upwards through the area of A as the vertical intensity Z sends downwards. In this case, the total number of tubes of force linked with A will be the same at the beginning and end of the turn, and the galvanometer will remain at rest. Hence, in terms of a constant (K) involving nothing but the area of A and the coefficient of mutual induction of the coils A and B, $Z = Ki_0$. With the instruments constructed hitherto, K has not been determined absolutely by measuring the geometrical configuration of the coils, but only relatively (by comparison with the value of Z deduced, according to the formula $H \tan I$, from measurements of H and I in non-polar regions).

D. W. Dye [30] designed an *absolute Z-magnetometer* which is a counterpart to the H-magnetometer of Schuster and Smith. A current i of known intensity is sent through a Helmholtz-coil system, with horizontal windings and vertical axis. The current is adjusted till its magnetic field at the centre of the coil just annuls Z; then $Z = ki$, where k depends only on the dimensions of the coil. The disappearance of the resultant vertical field is tested as follows, by the method of the vibration-galvanometer. At the centre of the system a small flat coil $(2 \times 1$ cm.), free to swing about a horizontal vibration-axis made of phosphor-bronze, is suspended in a frame which is adjusted so that the vibration-axis is alined along the magnetic meridian, and the plane of the vibration-coil is vertical. A comparatively strong alternating current, of frequency equal to that of the mechanical resonance-vibration of the coil, is led through the suspension-wire to the vibration-coil; hence this behaves like a small magnet with its axis perpendicular to the magnetic meridian, save that the direction of its axis is continually reversed. Hence it will begin to vibrate about the suspension under the action of the vertical magnetic field Z, unless Z is annulled exactly by the field of the Helmholtz-coil. L. F. Bates [28a] provides the resonance-coil with a core of μ-metal (1.17) of high permeability; an alternating current of a few milliamperes makes it a strong alternating magnet.

Observations with the Dye instrument at different times were found to be consistent with each other within about 0.3γ; the absolute value may be accurate to between 2 and 4γ. The absolute value depends, of course, on the accuracy of the international electric standards [105], which has not yet reached 10^{-5} of the total, and is therefore inferior to the accuracy of the standards of length and time.

2.6. Survey measurements on land. In *land surveys* the instruments used are generally similar to those employed for absolute measurements at observatories, except that the weight is reduced and the design is more robust, with a view to safe and convenient transport. The usual outfit consists of a theodolite, which is used both for astronomical work and for measuring D and H (Plate 6 *a, b* and Fig. 6).
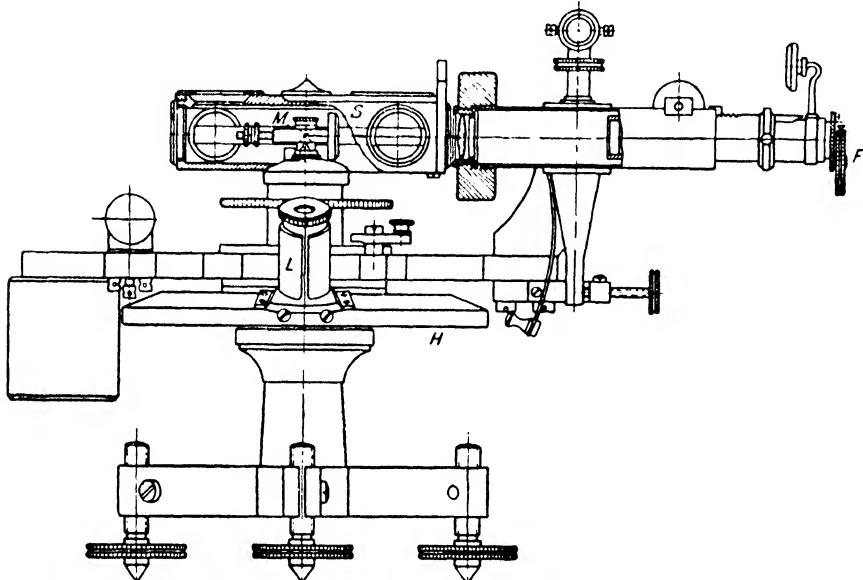
FIG. 6. Portable magnetic theodolite with a pivoted magnetic needle M, telescope F, mirror M, and eye-piece L to read the horizontal divided circle H. The same tripod and telescope are used for the deflexion experiment (see also Plate 6 *a*) and for the oscillation experiment (see Plate 6 *b*)

The needle is generally supported on a pivot, instead of being suspended by a thread. F. E. Smith designed a fluid suspension for the needle [29 *a*]; the magnet is attached to a float which is lifted by the fluid in the float chamber, and the pivot rests, with little pressure, against a jewel bearing (Fig. 7 and Plate 7).

The dip I is measured by a portable dip-inductor (to be fixed on the base of the theodolite, Plate 6 *c*), and, as in sea magnetic surveys, a string galvanometer (§ 7) may be used instead of a moving-coil galvanometer. Recently other electrical instruments have also been used for surveys.

For a general survey (a first-order survey, in the language of geodesy) the network of stations can be fairly wide, with an average distance of 20 to 40 km. Each chosen station must be representative for its neigh-
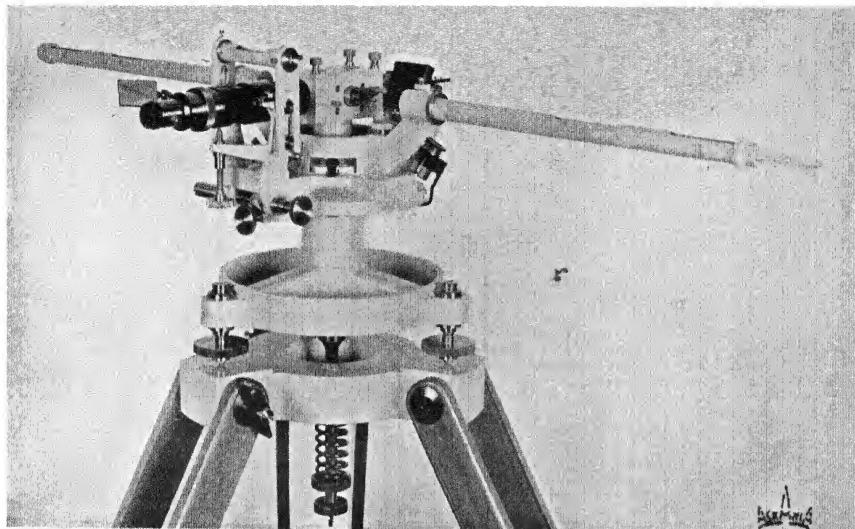
PLATE 6

(a)
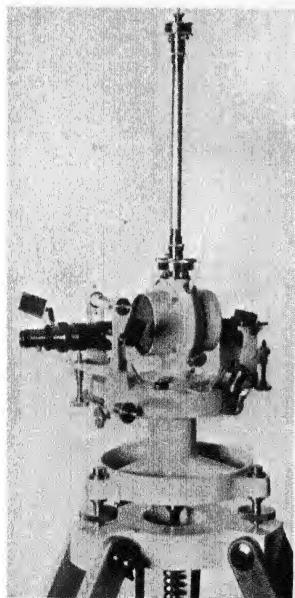
(b)
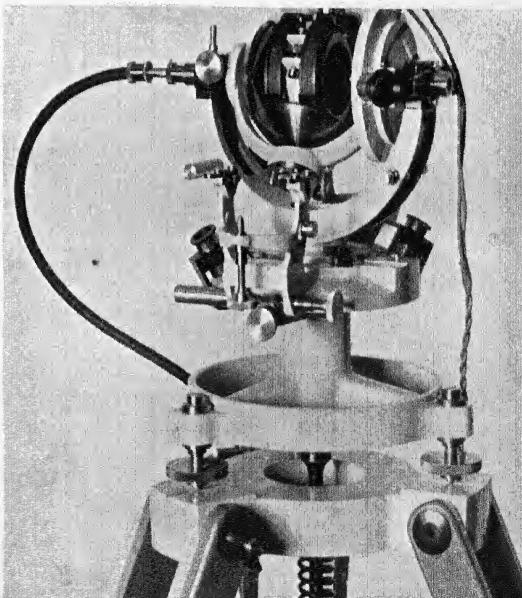
(c)

(a, b) A portable magnetic theodolite with a needle on a pivot. The same tripod is used for deflexion experiment (see (a) and Fig. 6) and, with the box, for oscillations (b)

(c) An earth-inductor mounted on a theodolite stand (Askania)

PLATE 7
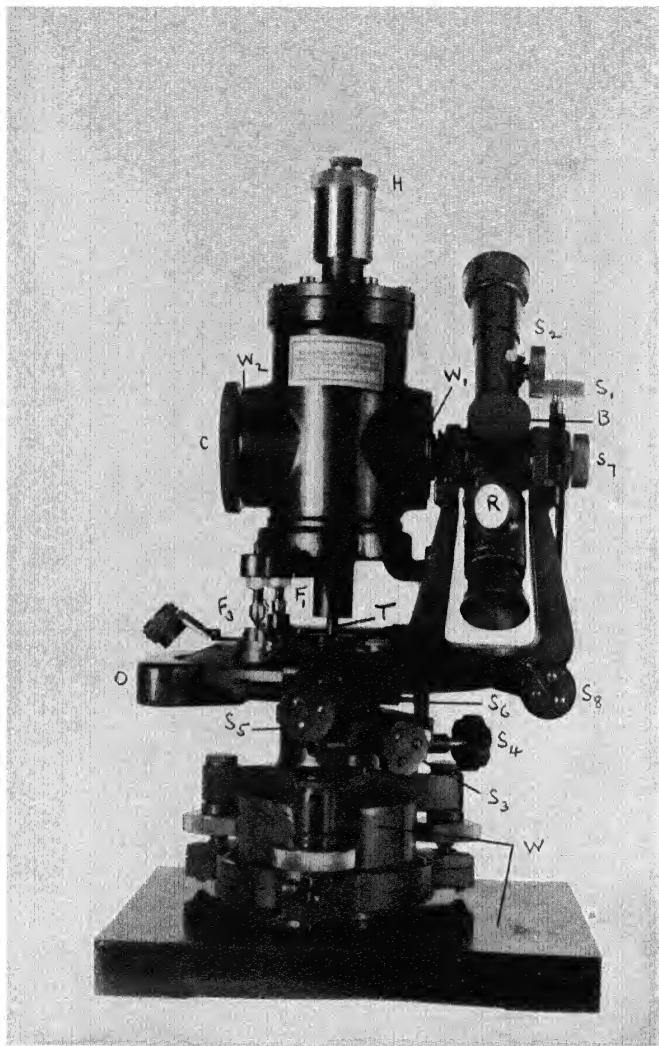
The Ordnance Survey magneto-theodolite (see Fig. 7)

bourhood; recent surveys therefore begin with a local survey, with local variometers (4.2), in order to select an undisturbed site for the station. This site is marked so that it can be reoccupied for repeat measurements. For instance, the Ordnance Survey, in the Magnetic Re-Surveys
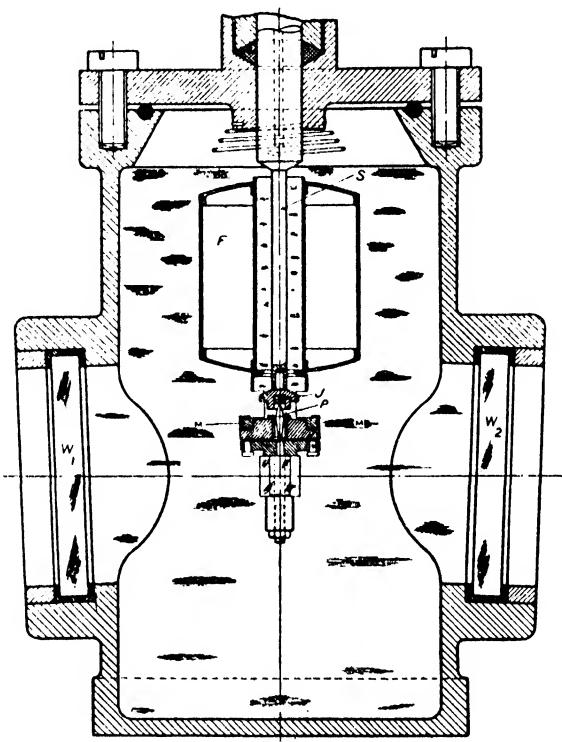
FIG. 7.　Float chamber and magnet system of the magneto-theodolite (Ordnance Survey): F float, P pivot, J jewel, mounted on the spindle S, which projects downwards through the centre of the float chamber; M, M two parallel cobalt steel magnets; windows W_1 and W_2 through which the position of the float is observed by means of reflections from one of the four sides of a glass cube (see also Plate 7)

of Great Britain between 1925 and 1932, made observations at 3 base stations and 172 field stations distributed over the whole country, including the surrounding islands. The survey observations are *reduced to a common epoch* (§ 8) by subtracting the time-variations intervening between that epoch and the time of each observation; the correction is determined from the records of one or more magnetic observatories (§ 9) not too far away.

Special relative instruments for *local surveys* (magnetic prospecting) will be described later (Ch. IV).

2.7. Survey measurements at sea. *At sea*, measurements are seriously hampered by the movements of the vessel and, in general, by the iron used in the construction. Instead of the ordinary needle, the *compass-card* is used for indicating the direction of magnetic fields. It consists of a light system of parallel horizontal magnets, turning about
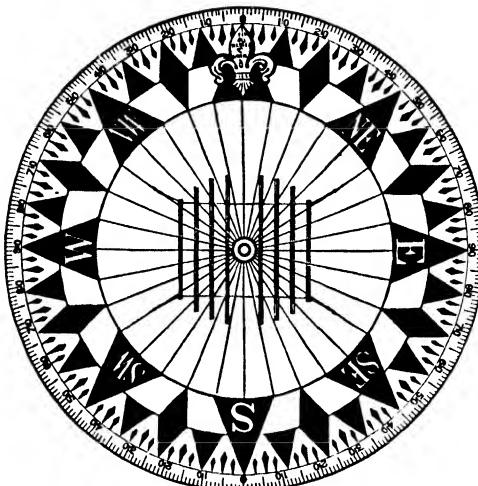
FIG. 8. Kelvin's compass-card

a pivot and bearing a circular card with a graduation at the border, so that the azimuth can be read relative to a mark on the case.

The famous compass-card devised by Kelvin (Fig. 8) in 1874 has a diameter of 25 cm. and weighs 13 gm. A small aluminium disk at the centre, attached to an inverted cup with sapphire crown resting on a sharp point, is connected by 32 stout silk threads with the aluminium 'card', the central part of which is cut away for lightness, so that the card is really only a rim; the threads also support a set of eight parallel needles of thin steel wire (lengths up to 8 cm.), with a total magnetic moment of about 200 Γ cm.[3] This card, owing to its great moment of inertia, has a free oscillation period of about 40 seconds, considerably longer than the chief oscillation periods of most ships. Kelvin saw that the 'perniciously great length of the compass-needles', 'becoming larger with the ships, by a process of Artificial Selection unguided by intelligence', was the reason for the unsatisfactory behaviour of earlier compasses [33–5].

Long experience in making compasses has resulted in their maintaining an admirably steady position even in rough weather; this is

achieved by increasing the moment of inertia of the card and by diminishing the friction on the pivot, particularly in the fluid compasses, filled with some alcoholic fluid. Disturbance due to the rolling of the ship is minimized by mounting the compass in gimbals (Plate 8 *a* shows a gimbal-stand designed for another magnetic instrument).

The influence which the iron in the body of the ship exerts on the declination indicated by the compass consists of three parts: a constant value representing the effect of the asymmetrical distribution of the iron, and two 'periodic terms' representing the influence of the permanent magnetization of the ship, and of the additional moment induced by the earth's field; the latter effect changes with the field. If D is the true value of the declination, D' the disturbed value, and ζ' the magnetic course (the angle between the compass-direction and the ship's axis), the *deviation* δ is, with sufficient approximation,

$$\delta = D - D' = A + (B_1 \sin \zeta' + B_2 \cos \zeta') + (C_1 \sin 2\zeta' + C_2 \cos 2\zeta'). \quad (13)$$

Treatises on navigation explain fully how the deviation is determined and removed by compensating devices [42–7].

Naturally the other magnetic elements are likewise disturbed by the ship. Hence ships constructed without the use of iron or other magnetic material are the most reliable for magnetic surveys, as was shown by the world cruises of the *Carnegie*, or the cruise of the non-magnetic yacht *Cecilie* in the Baltic.

In sea magnetic surveys the declination D is measured by the compass in conjunction with astronomical observations. The horizontal intensity H is found by deflecting the compass by magnets, using the principles of the magnetic theodolite.

Owing to its simple manipulation, the dip-circle has been retained longer at sea than elsewhere. It serves also for determining the total intensity F. This is done either by deflexion of the ordinary dip-needle in its vertical plane by magnets (just as H is measured by deflexion of the compass-needle in the horizontal plane), or by using a *loaded dip-needle*, i.e. a needle with a small non-magnetic weight fastened eccentrically to it. If I is the true dip and I' the 'loaded' dip, then the turning-moment of the earth's field on the loaded dip-needle is proportional to $F \sin(I - I')$, and that of gravity is proportional to $\cos I'$, so that, in equilibrium, $$F = C \cos I' / \sin(I - I');$$

C is an instrumental constant depending on the magnetic moment of the loaded needle and the distribution of its mass relative to its centre of motion.

Kelvin and Bidlingmaier pointed out that the movements of the ship not only cause the loaded dip-needle to swing but also produce a systematic change in its *average* position. This *dynamic deviation* affects all measurements at sea, and may spoil the results rather badly; it is proportional to the square of the eccentricity of the mass-centre of the swinging body. There is need for further theoretical and experimental investigations on it, such as those made by Peters [46].

The *marine earth-inductor* of the Carnegie Institution [36] is mounted, as usual for instruments on board, on a *gimbal-stand*. Plate 8 a shows a special gimbal-stand designed by Fleming. It consists of a suitable deck support with a U-shaped arm at the top. This arm carries a heavy supporting ring, in which two knife-edges are mounted diametrically opposite each other, and carry a second ring free to turn around the horizontal axis defined by the two knife-edges. The second ring carries, at the end of the horizontal diameter at right angles to the line of the knife-edges in the first ring, two knife-edges supporting a third (inner) ring, on which are provided suitable grooves for the footscrews of instruments. The second and third rings are carefully balanced, and the inner ring is properly loaded by a weight to ensure a condition of stable equilibrium.

The alternating current induced in the marine earth-inductor by the earth's field is not rectified, because of the difficulties in observing moving-coil galvanometers at sea; a string galvanometer is used instead. In this instrument the current is led through a thin silvered quartz string mounted between the poles of strong horseshoe magnets. Unless the axis of the inductor is parallel to the magnetic intensity F, the string will vibrate while the inductor revolves.

Bidlingmaier [37] developed the theory of the *double-compass* for measuring H at sea. This instrument consists of two similar horizontal compass-cards freely mounted with their centres at a vertical distance e above each other. If the earth's field were absent, the two cards would come to rest with the opposite poles above each other, but actually the horizontal intensity H tends to turn both cards into the direction of the magnetic meridian. The result is that, in equilibrium, the north–south lines on the cards will form an angle ψ with one another, each being turned out of the meridian, through $\frac{1}{2}\psi$, if both cards have the same moment M; the couple exerted by H on the upper card is then $MH \sin \frac{1}{2}\psi$, and that exerted by the lower card on the upper one is $(kM^2/e^3)\sin \psi$, the deflexion constant k being nearly unity. Hence

$$H = 2(kM/e^3)\cos \tfrac{1}{2}\psi.$$

PLATE 8
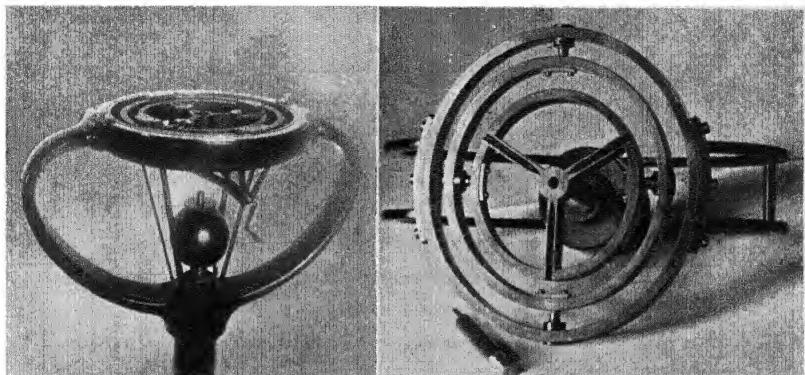
a. A gimbal-stand for a marine earth-inductor. View from the side (left-hand picture) and from above (on the right) (Carnegie Institution of Washington)
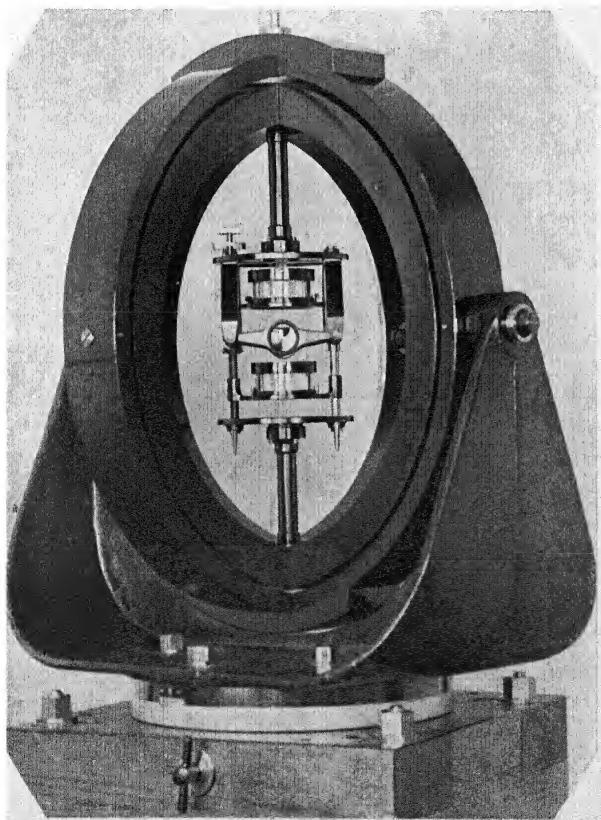
b. A compass-variometer mounted in a heavy gimbal-stand for marine use (Carnegie Institution of Washington, Model 4)

The compass-variometer of the Carnegie Institution of Washington [40] is based on the same principle (Plate 8*b*). The two compass-cards are replaced by two circular steel disks (diameter 24 mm., thickness 0·3 mm.), permanently magnetized along a diameter by a coil arrangement; each disk is mounted in its own cell, provided with liquid damping. These cells are placed vertically above each other, so that their distance can be changed and read accurately. The upper disk-magnet is graduated on its lower surface, and the lower magnet on its upper surface; by a prism arrangement placed between the magnets, their relative angle can be read. The period of oscillation of the magnets is of the order of 2 seconds. The gimbal system has a long period (40 seconds); it consists of two rings rigidly connected, mutually perpendicular, mounted vertically, and movable about a horizontal axis supported by a yoke and lying in the central vertical plane of the supporting ring, and perpendicular to the plane of the other (the inertia ring). The total weight of the gimbal system shown in Plate 8 is 450 pounds; the centre of gravity of the moving-ring system is displaced only 0·3 mm. below the horizontal axis. The instrument is designed for local surveys of anomalies at sea.

A few attempts have been made in balloons and dirigibles to measure the earth's field at heights of some kilometres above the ground. Theory shows (18.10) that the magnetic elements at such heights are likely to be so little different from their ground values that the differences will not be determinable in view of the difficulties of measurement in the free air. But the geomagnetic field is used in aeronautics e.g. in the earth-inductor compass driven by a small anemometer and indicating the direction by vanishing of the induced current [48*b, c*]; dynamic deviation in aircraft causes the 'northerly turning error' [48*a*]. At heights of 200 or 300 km., however, the change of the magnetic intensity is a few per cent., and this has been detected by radio methods (15.5).

2.8. The secular magnetic variation. Absolute observations determine the value of F at a given point and time. Later observations at the same point usually reveal a change in F. The absolute observations made regularly at magnetic observatories enable the changes in F at these points to be followed in some detail. They are found to vary considerably from one observatory to another, so much so that, except over short intervals of time, it is unsafe to infer by interpolation what are the changes in F at points between widely separated observatories. Consequently magnetic surveys have to be repeated at intervals of several years if our knowledge of the distribution of F over the earth

is to be kept up to date. For this purpose 'repeat stations' are selected, with special care regarding the absence of natural or artificial disturbance.

By such repetitions of surveys it has been found that the earth's field F undergoes a large variation of complicated type, which is called the *secular* magnetic variation. It often proceeds in the same direction for a long period of time at any one station.

Where there is no magnetic observatory our knowledge of F depends on occasional absolute observations, generally made in the course of a systematic survey. It is customary to publish the actual measurements of H, D, and I, and the times and places at which they were taken (often with topographical details to assist in finding the same station, in case it is to be reoccupied later).

Furthermore, the single observations, taken in the course of surveys at different times, are sometimes *reduced to a common epoch*; this epoch is taken the same for all the stations, and is generally chosen near the middle of the period occupied by the survey. The reduction may be based on the assumption, sufficiently warranted in non-polar regions, that within a distance of about 500 km. from an observatory the temporary changes of F are nearly the same everywhere, except during marked magnetic disturbance. Therefore the difference $\Delta F = F - F_0$ between the value F at some point P of the survey, and the simultaneous value F_0 at the observatory, will be nearly a constant characteristic of P. If F_{0e} is the value for the observatory at the epoch chosen, the simultaneous value at P is given by $F_e = F_{0e} + \Delta F$.

Alternatively the reduction may be to the annual mean value of the element for a year centred at the chosen epoch; this is the more appropriate plan when it is desired to obtain the secular variation at the station, in the interval between two epochs. In regions where there is no magnetic observatory, the reduction is made by interpolation, and is based only on data for the secular variation derived from relatively scarce repeat-measurements.

2.9. Continuous magnetic registration. Besides the secular variation, other changes in the earth's field, partly periodic and partly irregular, are continually in progress. These also it is the function of observatories to record. A continuous register of the changes in F at an observatory is obtained by means of instruments called *variometers*, which provide curves called *magnetograms*, showing the variations in each of three elements (usually D, H, and Z, but sometimes X, Y, and Z). The variometers are designed only to register the continuous

changes in the elements, on a consistent plan. In conjunction with the absolute observations of *F* made at suitable intervals, they enable the value of *F* to be found at any intervening time. A separate variometer is required for each element registered.

The changes in the magnetic field are indicated by proportional changes of direction of a needle suitably disposed. These movements are magnified and *registered* by *photography*; it is sufficient to describe
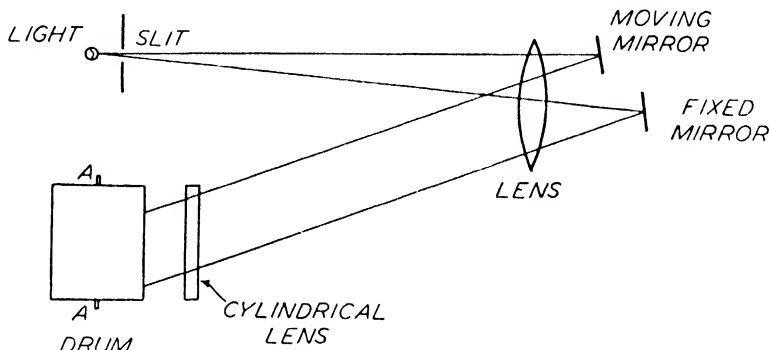
FIG. 9. Schematic plan of a photographic recorder

the method in detail in the case of the *D*-variometer (Fig. 9). This consists simply of a horizontal magnetic needle free to follow the changes in the horizontal direction of *F*. The source of light is a small lamp, in front of which is placed a screen with a narrow slit; the slit may be dispensed with if the lamp has a single straight filament. The light passes from the lamp through a lens to a mirror attached to the magnet, and is reflected again through the lens. The distance of the slit or single filament from the lens corresponds to the focal distance. The reflected beam is focused on the surface of a suitably disposed cylinder (or drum), with horizontal axis, after being contracted to a spot by means of a horizontal cylindrical lens of short focal distance placed in front of the drum. The motions of the needle and mirror cause this spot to travel to and fro along a generator on the cylinder. Owing to the large distance from the mirror to the drum, small displacements of the spot are proportional to the deflexions of the needle; the spot moves just like the end of a needle of the length of the light path from the lamp to the cylinder. By clockwork, the drum is made to rotate with uniform speed. This regular motion of the cylinder about its axis, combined with the motion of the spot of light, registers a curve on a sheet of photographic paper wrapped round the cylinder.

It is convenient to use the word *magnetograph* to signify a magnetic
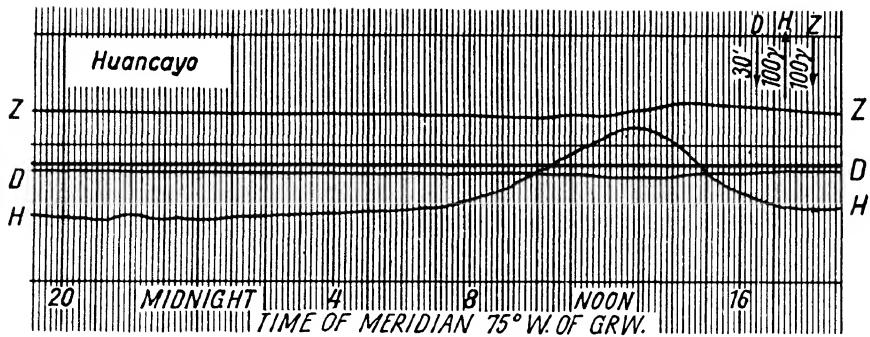
Fig. 10 a
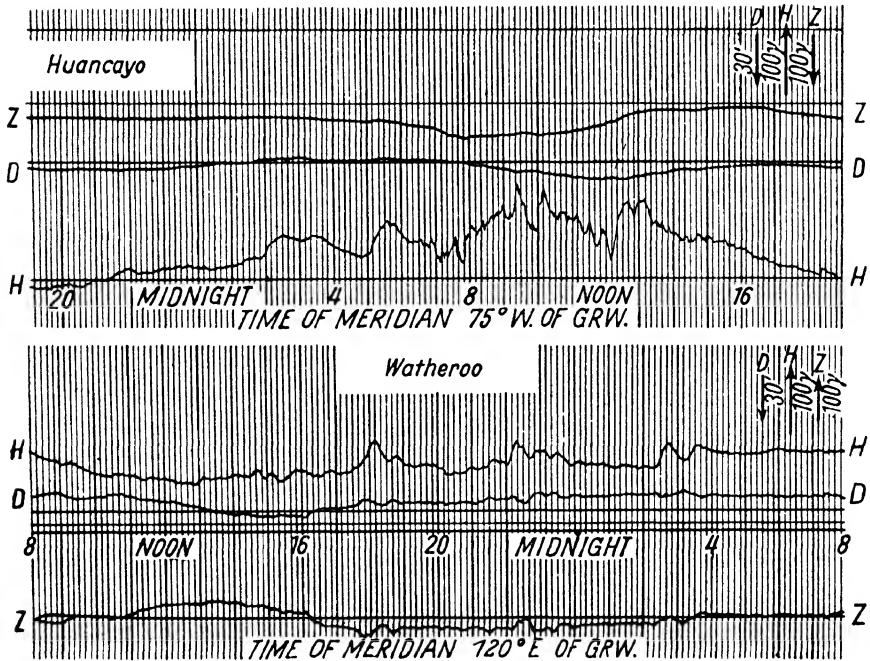
Fig. 10 b

variometer together with the recording mechanism by which permanent records of the magnetic changes are obtained, these records being termed *magnetograms* (cf. *telegraph* and *telegram*). The variometers when used merely for eye-readings of the magnetic element would thus not be called magnetographs. It should be added, however, that in geo-
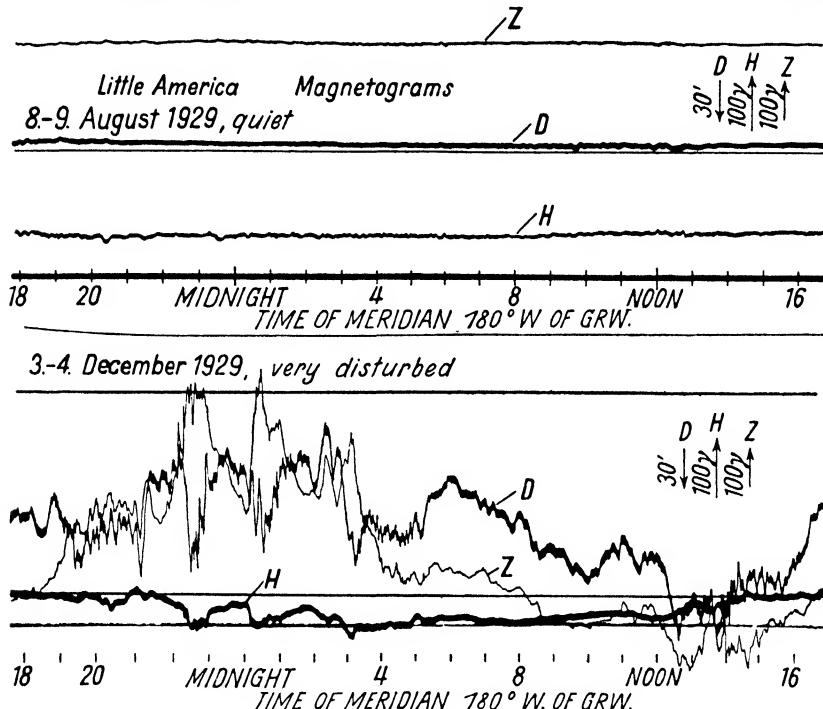
FIG. 10 *a–c*. Magnetograms for a quiet day (1929 August 9) and a disturbed day (1929 December 4), from Huancayo (Peru), Watheroo (W. Australia), and Little America (Antarctic). The selected intervals begin and end at Greenwich midnight; the local mean time is indicated for each station

magnetic literature the word magnetograph is often used to signify the records, here called magnetograms.

Reduced reproductions of such curves are shown in Fig. 10 *a, b, c*. The original sheets are about 40 cm. long and 20 cm. broad; the vertical lines (time-marks) indicate the direction of the generators of the cylinder from which the sheet was unwrapped, and they mark off hourly and ten-minute intervals of time, measured along the transverse direction; the whole length of the sheet represents 24 hours, the time-scale on the original being 15 mm. per hour. A straight line, called a *base-line*, from which ordinates can be measured, is also traced photographically

on the sheet; it is generated by a spot of light reflected from a fixed mirror fastened to the frame of the variometer. Other curves record the temperatures of the variometers.

Photographic recording has the disadvantage that the curves remain invisible until the sheets are developed. *Direct recording* would be desirable, if only in order to enable an observer to detect at once that a magnetic disturbance is in progress. The magnetic force variations are, however, too weak to work mechanical recorders, although a few attempts have been made to construct direct recorders [66]. The mechanical difficulties can be avoided by using a *photo-electric* record. In this the light from the mirror falls, in a broad band, on two photo-electric cells placed side by side. The difference between the currents produced in the two cells is recorded by a galvanometer; it is zero if the band is symmetrical, and positive or negative if the band covers more of the area of one cell than of the other [4.13 c].

Plate 9a shows a photographic recording apparatus which also permits eye-readings to be taken.

FIG. 11. Unifilar *H*-variometer: T torsion-head, L lens, S mirror attached to the magnet M moving in a shallow copper box for electromagnetic damping; B fixed base-line mirror (Askania)

The needle of the *H*-variometer (Fig. 11) hangs transverse to the magnetic meridian, from which it is deflected either by twisting the suspension (which may be unifilar or bifilar), or by the field of auxiliary magnets. In this position its direction is unaffected by small changes in *D*, but it varies through small angles proportional to small changes in *H*, the sensitivity being suitably adjustable. In the same way, *X*- or *Y*-variometers contain needles kept approximately in the east–west or north–south direction.

The *Z-variometer* contains a horizontal magnet which can turn about a horizontal axis lying in the magnetic meridian. In this position it is unaffected by small changes in *D* and *H*, but its inclination alters by small amounts proportional to changes in *Z*. To keep the needle nearly

PLATE 9
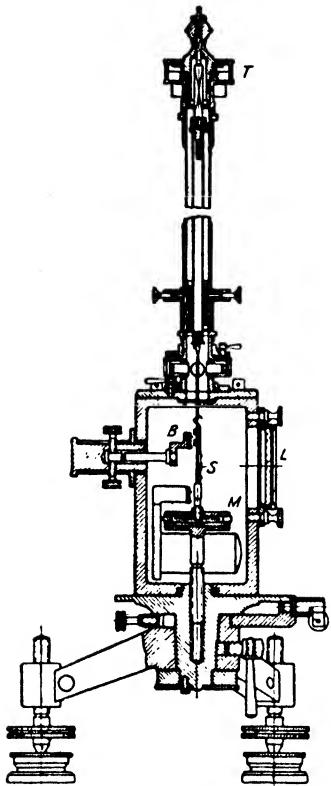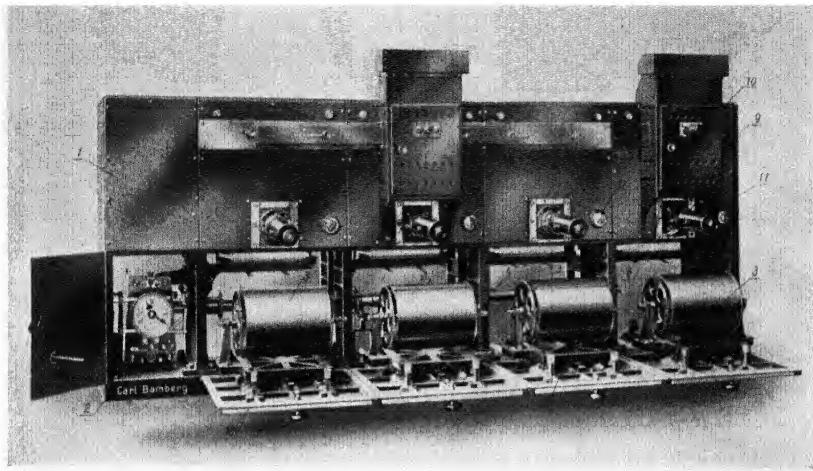
a. Large magnetic recording apparatus with four drums and with telescopes for eye-reading, as used at Niemegk Observatory (Askania)
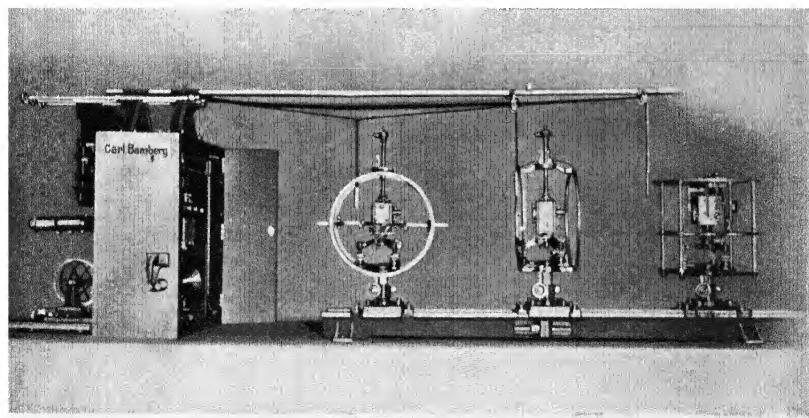
b. On the right, magnetic variometers for *H*, *D*, and *Z*, and on the left, the recording apparatus (Askania)

*

horizontal, the horizontal distance d between its axis of rotation and its centre of gravity must, by means of small screws or otherwise, be adjusted once for all so that the turning-moments of gravity and of the mean value of Z compensate each other; d is very small, never greater than about $0 \cdot 1$ cm., so that the equilibrium in Z-variometers is of a delicate nature. Usually the needle rests on an edge made of agate, as in the *magnetic balance* designed by Lloyd; in some instruments it is suspended on two taut horizontal quartz-threads lying north and south, and twisted so as to keep the needle horizontal, and perpendicular to the magnetic meridian. The mirror is fixed with its normal vertical; glass prisms reflect the beam of light into the horizontal direction.

Care must be taken to arrange the different variometers in the same room so that their magnetic fields do not disturb each other (Plate 9 b).

In the D-variometer the needle turns freely as D changes, but in the H- and Z-variometers the needles are twisted by couples which are proportional not only to the changes in H or Z but also to the magnetic moment of the needle. This, as also the dimensions of the instrument, may change with temperature, and so the temperature τ of the variometers must also be registered. All readings are reduced to a standard temperature τ_0 by means of a *temperature coefficient* ϵ, which is determined from time to time by heating or cooling the room, or a smaller enclosed space round the instrument.

If fixed bar-magnets are used in maintaining the equilibrium of the needle, their fields also will change with temperature, so that by careful adjustment it is possible to *compensate* the readings for temperature. But there remains an uncertainty, for this compensation fails when the temperature is not sufficiently uniform in time and space. Therefore steps are taken in many modern observatories to prevent any changes of temperature which might impair the accuracy of the observations, by housing the variometers underground or in double-walled chambers, often fitted with artificial heating controlled by thermostats.

The procedure required in order to obtain exact values of the magnetic elements at any instant, from the absolute observations and the magnetograms, will be briefly outlined here. Some details of the various refinements of technique will be given in part B of this chapter; descriptions may also be found in observatory publications [52].

The value S of the magnetic element, for an instant at which the ordinate of the magnetogram is a mm. measured from the base-line, is given by

$$S = S_0 + ea + \epsilon(\tau - \tau_0), \tag{14}$$

the last term being the temperature correction; S_0, the value corresponding to $a = 0$ (and $\tau = \tau_0$), is called the *base-line value*, and the coefficient e is called the *scale-value*. For temperature-compensated instruments $\epsilon = 0$, and for temperature-controlled instruments we can take $\tau = \tau_0$, so that $S = S_0 + ea$.

The scale-value of the D-variometer is calculated from the length of the light-beam from the magnet-mirror to the recording drum. For the other variometers, e is determined by occasional special experiments in which the change of a is recorded when the needle is deflected by a known artificial change of the field by means of magnets or, better, by Helmholtz coils (§ 5), as shown in Plate 9 b.

The base-line value S_0 is determined by means of the absolute observations, which determine S at a known time, so that the corresponding a can be read from the magnetogram. Since absolute observations require a certain time, a must be a mean ordinate over this time; better still, each reading of the absolute instrument is associated with the reading of a for that instant.

Generally the base-value S_0 changes slowly, owing partly to the loss of moment of the magnets, and partly to slight mechanical changes. The small changes of S_0 found at successive absolute observations are smoothed and distributed uniformly over the time between. Abrupt changes of S_0 are indicated by jumps in the curve. They may arise in the Z-balance by minute changes in the distribution of mass, and must be treated individually.

The scale-values ordinarily used are $e = 1'$ per mm. in D, and 2γ to 5γ per mm. for force components. It is easy to attain higher sensitivity, but not advisable for ordinary purposes, because the gain in the accuracy of the readings is offset by uncertainties inherent in the variometers, unless several recording sets with different sensitivities are set up.

During great or rapid magnetic disturbance, loss of photographic trace may occur for two reasons. The spot of light may move so fast that no trace, or only a confused fragmentary trace, is left; or the spot may go beyond the limits of the photographic paper. To provide against the latter, additional mirrors or prisms, placed in the light-path, are sometimes used, which give additional light-spots separated from the central one by a little less than the width of the record. This device, however, does not obviate the loss of trace caused by rapid motion of the light-spot. Therefore it is advisable that magnetic observatories should set up special insensitive instruments (*wide-range recorders*) in addition to the ordinary variometers; the scale-values of

the insensitive recorders must be chosen so that even the greatest and most rapid variations (say $3,000\gamma$ to either side of the normal values) will be recorded; 30γ per mm. is a suitable scale for this purpose. Such insensitive sets do not need so much supervision as those of ordinary sensitivity, and their records are only used—perhaps once in ten years —when the ordinary records fail in a great magnetic storm; but they should be introduced at more ob- servatories than hitherto, since a loss of trace during a great magnetic disturbance—that is, during some of the magnetically most interesting periods—should certainly be avoided.

An optical device due to Schmidt [52] provides sensitive and insensitive records from the same instrument. Fig. 12 shows, for instance, the mounting for the mirrors of a vario- meter for declination or horizontal intensity. In addition to the ordinary vertical mirror I, which gives the curve of ordinary scale, a second in- clined mirror II is fastened to the magnet (its weight is balanced by a counter-weight). A fixed mirror III, attached to the frame of the instru- ment, reflects the light from the lamp

FIBRE

COUNTER-
WEIGHT

INCLINED MIRROR-MOUNTING
AFTER ADOLF SCHMIDT

FIG. 12. Schmidt's device to provide two records of different sensitivity from the same magnet

on to the mirror II and back to the photographic paper; this second spot gives the insensitive curve. If α is the angle between the mirrors I and II, the deflexions of the two light-spots are in the ratio $1:\cos\alpha$. This *dual-sensitivity* device could probably find useful application in many other types of physical recording instruments.

At stations near the magnetic poles the horizontal components vary by amounts that are of the same order of magnitude as their mean values; the ordinary method of registration can there be maintained only by introducing strong artificial fields by means of auxiliary mag- nets, so that the sensitivity is kept low (i.e. the scale-value high)— otherwise frequent loss of trace occurs. Lower scale-values are used, however, in the quick-running magnetographs described in § 10.

Of the three variometers, that for D is the simplest and most accurate

(changes of temperature have no influence on its registrations), and that for Z is the most difficult and least reliable. In all the three elements, however, it is easier to measure *changes* (to 1γ, say) than to determine the absolute values (or base-values) with the same accuracy.

The inferred value of an element at any particular instant is subject to an error, additional to those of the kinds already indicated, depending on the identification of the point on the magnetogram which corresponds to the instant; this error will be the greater, the more rapid the variation of the element at the time. Sometimes the time is marked on the magnetograms by short blank spaces in the curves, caused by shutting off the light automatically at known times; another method is to use black lines passing right across the sheet, due to light, focused by the cylindrical lens, from lamps which flash up every hour or more frequently. With the ordinary time-scale (15 or 20 mm. per hour) it is difficult to determine the time corresponding to any assigned point on a magnetogram, or the point corresponding to an assigned time, more accurately than to one minute. The recording drums are now often designed so that for special researches, in which more accurate timing is required (for instance, during absolute observations), the cylinders can be rotated at ten or twelve times their usual rate, thus increasing the time-scale on the magnetograms in the same ratio. Naturally such quick-running magnetographs use much more photographic paper than ordinary magnetographs, if the paper is changed after each revolution of the drum.

2.10. Quick-run magnetographs of Schmidt and Là Cour. Schmidt [71] invented a valuable device to save photographic paper in quick-running magnetographs. The lamp is placed at the focus of a big cylindrical mirror; this reflects a parallel beam of light, which passes through many equidistant slits in a screen, instead of through one slit only, as in the usual arrangement. The plane mirror attached to the magnet reflects the beam of light coming from each slit, so as to fall on the cylindrical lens of the recording drum. By passing through this lens it is focused into a light-spot on the surface of the drum. The series of slits produce a series of equidistant light-spots ranged along a generator of the cylindrical drum. As the drum rotates, the spots, which all move in unison to or fro along the generator, would produce a series of parallel curves on the photographic sheet, were it not that all but one (or at most two) of the spots are screened from the sheet by means of a diaphragm placed in front of the drum, the opening of the diaphragm being only slightly wider than the distance between adjacent light-spots. Thus only one spot at a time is able to register

PLATE 10
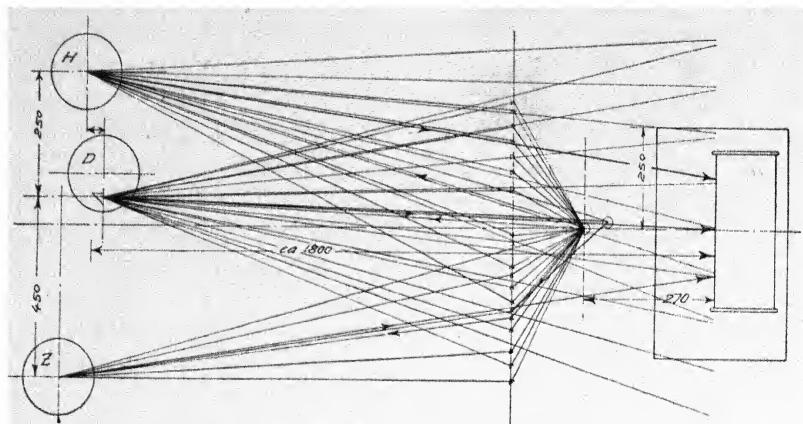
a. Schematic ground-plan of three magnetic variometers with the La Cour quick-run recorder, showing some of the light-paths between the lamps, prisms, variometers, and recording drum
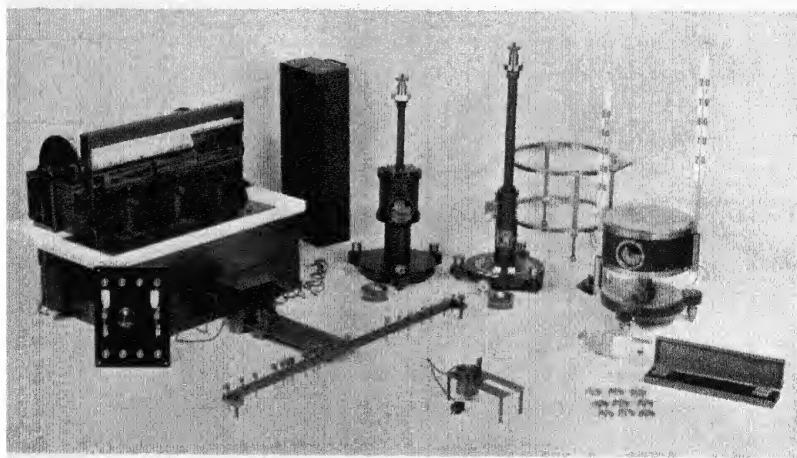
b. On the left, the La Cour quick-run recorder with its cover removed, and on the right the three variometers

on the cylinder, except when the spot approaches the limit of opening, when a second spot comes within the opening near the other limit (Fig. 13). The diaphragm moves sideways by its own width in the course of one revolution of the drum.

La Cour [72] modified Schmidt's device by using a straight filament lamp and a row of small prisms placed before the lamp (Plate 10), which likewise produce a series of images of the filament, only one of which
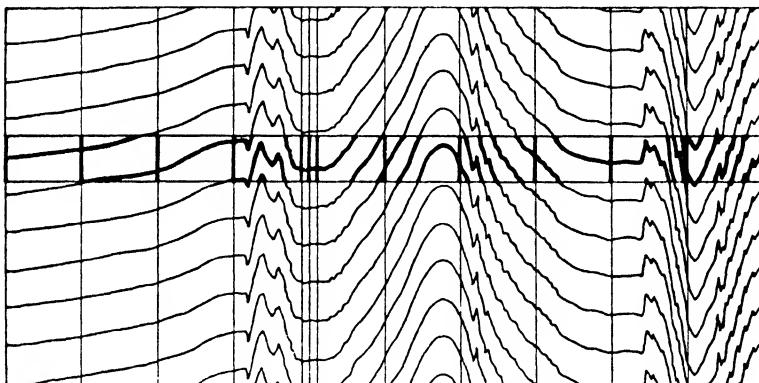
FIG. 13. To illustrate Schmidt's method for economical quick-run registration

is allowed to fall on the drum. In the magnetographs which he designed for use during the Second International Polar Year (§ 17) he used forty-four such prisms, giving light-spots separated by distances of about 7 mm. These were screened by a moving diaphragm having an opening of about 8 mm. He arranged that all three variometers (D, H, Z) should record during 24 hours on the same drum, each record extending over a third of the width of the drum (Fig. 14). The paper used is of size 30×39 cm., and the time-scale is 3 mm./minute.

2.11. Measurements of rate of time-change. In order to reduce friction and to shorten the time of oscillation of the needles in magnetic variometers, their magnetic moment must be as large as possible for the chosen size and mass; cobalt steel is very satisfactory in this respect. But the observation of rapid changes of F with the usual variometers is limited by the moment of inertia of the needles, since their free periods cannot be reduced much below a second. Therefore the *rate of change* of a magnetic element, for example dZ/dt, is best measured directly by the current which the varying field induces in a suitably disposed circuit containing a galvanometer. To register dZ/dt the circuit should lie in a horizontal plane; such an arrangement has been or is in use at various observatories (such as Eskdalemuir, Tromsö, Val

Joyeux, and Watheroo). The records thus obtained show interesting details not easily visible on the ordinary magnetograms, especially at the commencement of magnetic storms (Fig. 15) and during micropulsations (Ch. X).

Such a cable, of 12 km. length, was first laid out by Giese, at the suggestion of von Siemens, at the German station Kingua Fjord (Baffin
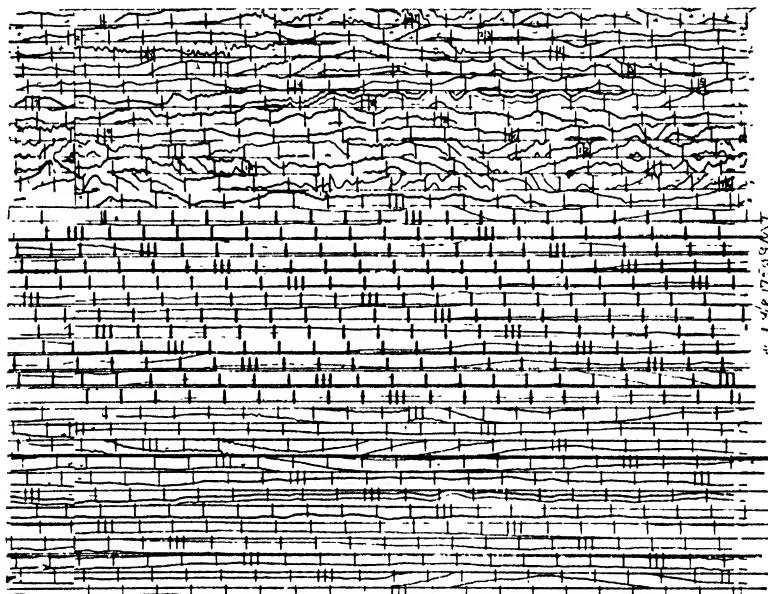
FIG. 14. A sample record for 24 hours, and for all three magnetic elements, from a La Cour quick-run recorder

Island) in the Polar Year 1882–3, but the arrangement proved too sensitive. Ebert [73] used a cable of fifteen strands connected in series and wound around an area of 2,000 square metres, so that the total area of the windings was 3×10^8 cm.^{2} From 1 (43), it follows that from such a circuit we obtain a total induced electromotive force $E = 3 \times 10^{-5}$ volts for $dZ/dt = 1\gamma$/second, or, with a total resistance of 150 ohms, a current of 2×10^{-7} amp. This can easily be measured by a string-galvanometer.

By suitable damping of the galvanometer to the limit of aperiodicity, and by making the resistance of the loop itself small, Rössiger [75] was able to obtain records showing the rapid changes of Z instead of dZ/dt. With multiple reflections from the mirror of the galvanometer (giving an increase of the optical path) he obtained a sensitivity of 0.05γ per mm.

deflexion on the photographic sheet, for periods of less than 10 seconds; slower changes are, of course, shown on a reduced scale.

Aschenbrenner and Goubau [76] have introduced a method of recording small and rapid geomagnetic variations, which is based on quite different principles. It depends on the fact that the magnetic induction B induced in iron by a field of intensity H is not proportional to H
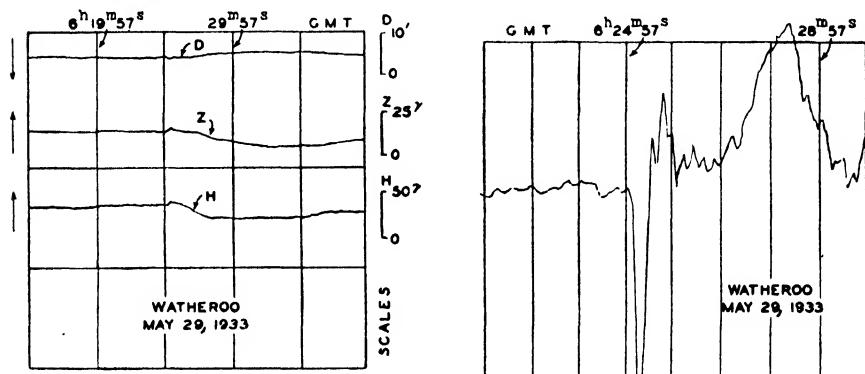
FIG. 15. A sudden change in the magnetic elements, recorded at Watheroo (W. Australia). *Left*: the quick-run record; the vertical lines are time-marks at five-minute intervals. *Right*: the record made by a Crichton Mitchell vertical-intensity inductometer, showing dZ/dt; the time-marks are at one-minute intervals

(except when H is small); this can be represented by supposing that the relation between B and H is approximately

$$B = aH - bH^3,$$

where a and b are positive constants (there is no term in H^2 if, as we suppose, the coercive force is small), so that B is reversed with H. Imagine two similar parallel iron bars, on which coils carrying a current $\iota = \iota_0 \sin \omega t$ are wound in opposite directions, so that the alternating field in the bars is $\pm H_0 \sin \omega t$. Let H_1 be the component of the geomagnetic field along the bars. Then the magnetizing fields are $H_1 \pm H_0 \sin \omega t$, and the corresponding values of B are

$$a(H_1 \pm H_0 \sin \omega t) - b(H_1 \pm H_0 \sin \omega t)^3.$$

If a secondary coil is wound round the outside of the two bars, the total induction through it will be proportional to the sum of these two values of B, namely to

$$2\{aH_1 - b(H_1^3 + 3H_1 H_0^2 \sin^2\omega t)\},$$

so that it contains a constant term, and a periodic term proportional to $H_1 H_0^2 \cos 2\omega t$ (since $\sin^2\omega t = \frac{1}{2} - \frac{1}{2}\cos 2\omega t$). Hence an alternating

current with period π/ω will be induced in the secondary coil, and its amplitude will indicate the value of H_1 if H_0 is known. If H_1 varies, the changing amplitude of the induced current will indicate the variations of H_1, so long as these are slow compared with the alternations of $\cos 2\omega t$. In practice the two parallel bars are replaced by a ring (as if the ends of the bars were joined by curved pieces) and the induction is increased by the use of μ-metal; also the induced currents have components proportional to $\cos 3\omega t,\ldots$, because the above relation between B and H is only approximate; but it is easy to 'filter out' the one harmonic in $\cos 2\omega t$, for measurement. By this method rapid variations of H_1 as small as $0\cdot 3\gamma$ have been observed.

2.12. The accuracy of magnetic observations. The standard of absolute accuracy aimed at in geomagnetic observations is 1γ. The accuracy actually attained naturally differs from one observatory and also from one element to another, and during the period, of about a century, since magnetic observatories have been established, there has been a gradual advance in the methods and instruments used. The reliance to be placed on any particular set of observations can be gauged only by a careful examination of the methods by which they were derived, or from comparisons in which world-wide and local effects are separated, as was done by La Cour [77], in a valuable critical discussion of European observatory data. Similar remarks apply to the accuracy of magnetic surveys.

The error in any component X, Y, and Z will depend on the separate errors in H, D, and I, according to the equations 1 (10) and (11). Since H is usually less than $0\cdot 4\ \Gamma$, an error in D of $0\cdot 1'$ (or $0\cdot 1/3438$ in angular measure) represents an error in X or Y not exceeding about 1γ: an error of $0\cdot 1'$ in I produces an error in Z which increases from about 1γ at the equator to infinity at the *poles of magnetic dip* (that is, the points where $I = 90°$).

2.13. Time-reckoning in magnetic observatories. Various systems of time-reckoning have been employed at magnetic observatories, and a brief outline of the points involved is desirable.

The true or *apparent* local solar time t' at any station P is a measure of the westward angle from the meridian of P to the *midnight* meridian (opposite to the *noon* meridian, or meridian half-plane, which passes through the sun). This angle is C'NP in Fig. 16, which shows the earth viewed along the geographical axis, from a distant point above the north pole N. For stations on the meridian of Greenwich (G) the time is called Greenwich apparent solar time t_0'. If the longitude of P east

of Greenwich is λ, reckoned in time units at the rate of 1 hour per 15°, then (cf. Fig. 16)

$$t' = \lambda + t_0'. \tag{15}$$

Owing to the varying rate of orbital motion of the earth, the solar day, or the earth-rotation period relative to the meridian plane through the sun, varies in duration in the course of the year. Hence the hours of apparent time are not constant. *Mean* solar time replaces this variable reckoning by a uniform one, with the same zero-point at the vernal equinox (the epoch when the sun is crossing the plane of the earth's equator from south to north), and the same total measure in the course of a year. The difference between Greenwich or local mean time, denoted by t_0 and t, and the corresponding apparent times, t_0' or t', is called the equation of time, e, so that

$$t_0 = t_0' + e, \qquad t = t' + e, \qquad t = \lambda + t_0. \tag{16}$$

The value of e at the middle of each month is given in angular measure in § 15.
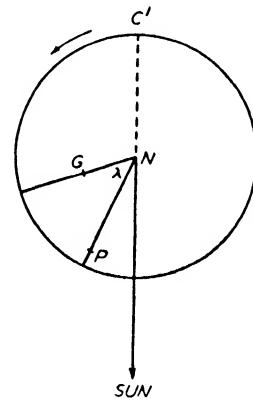
Fig. 16. The earth viewed from above the north pole N

Owing to the great influence of Gauss upon the study of terrestrial magnetism, many magnetic observatories, such as Greenwich, once employed Göttingen local mean time, partly in order that hourly readings of the magnetic elements (which in the early days were taken by eye) might be compared with simultaneous readings made at his observatory at Göttingen. In many cases the hours of observation were afterwards changed to those of local mean time at the particular observatory. Still more recently there has been a tendency to adopt standard mean time, differing from Greenwich mean time by a whole number of hours, so as to approximate to local mean time as far as this limitation permits. A few observatories not in or near the Greenwich meridian now tabulate their hourly data according to Greenwich mean time instead of according to the standard time of their zone.

The interval of 24 hours of mean solar time succeeding each epoch of mean midnight (0^h) at Greenwich is called a Greenwich mean day, as distinct from the local mean day at any station P, reckoned from one local midnight to the next. Local mean days for stations in different longitudes clearly cannot coincide.

2.14. Data published by magnetic observatories. Many of the principal magnetic observatories publish annual tables of hourly values of the three elements which they register by variometers, usually either

D, H, and Z or X, Y, and Z. These values may refer to the exact hours of the time-reckoning used (§ 13), or they may be means for hourly *intervals*, either centred at, or (preferably) extending between, the exact hours. Hourly means use the records more fully than do the instantaneous hourly values, and also eliminate some of the accidental character attaching to instantaneous values owing to chance disturbances; they were introduced by Schmidt in 1905 [G 49].

Some observatories also publish the extreme or 'absolute' daily range (R) of each element during each day, and/or the actual daily maximum and minimum values, with the times at which they occur. Others publish the daily range (R') in the hourly values of each element; naturally R' is in general less than R. This information, with the hourly values, sufficiently indicates the magnetic changes on days when the magnetograms are smooth and simple in form (cf. Fig. 10 a), but not on days when the variations are rapid or irregular (Fig. 10 b). Some observatories publish supplementary notes or descriptions of the curves for such days; when the changes are too complicated to be briefly described, reproductions of the actual curves are printed, for which a time-scale of 15 mm./hour, and ordinate-scales of $1'$/mm. for D and 5γ/mm. for H and Z, have been internationally adopted (see §§ 20 and 6.13). The dates of days for which reproductions are desired are published along with the international magnetic character-figures (6.2).

In addition, observatory publications usually give further results summarizing certain features of the above data. These include mean values of each element, and sometimes of other elements derived from them, for the whole year, for the twelve calendar months, and for each individual day (i.e. annual, monthly, and daily means); in the last case the day may be a Greenwich day or a day beginning at 0^h of the system of time-reckoning used at the observatory. The choice of the Greenwich day is preferable in that it gives information about the earth's mean field $\bar{F}_i$ (1.6) for a definite interval i (= a Greenwich day), the same everywhere. The twelve monthly means indicate the broad features of the annual variation of F; successive yearly means indicate the general course of the secular variation (see 6.11).

Observatory publications usually give, for each year and for each calendar month, the mean values of each element at each individual hour, for all days in the month, and/or for one or more groups of selected days. Such groups now usually include one comprising the five *Greenwich* days during the month which are internationally chosen as being the quietest, and also a similar group of five international most

disturbed days (6.2). The annual or monthly mean hourly values indicate the mean daily variation of each element during the year or month, on all or on selected days. The monthly mean hourly values for all days often refer to the day starting with 0^h of the local or standard time-reckoning, instead of to Greenwich days.

Sometimes the mean for all the days of a group is subtracted from the corresponding sequence of 24 monthly mean hourly values; the resulting sequence is called a *mean daily inequality*; it gives for each hour the mean hourly departure, in the group of days to which it refers, from the mean for all hours. Some observatories which register D, H, and Z publish monthly mean hourly values or daily inequalities for X and Y also, derived by computation from those for D and H.

Mean daily inequalities may refer to the hours 0 to 23 or 1 to 24, and for any given group of days it is usually found that the mean values for 0^h and 24^h are not quite equal, owing to the fact that the values at the two epochs of midnight bounding the individual days are unequal. The change from one hour (usually midnight) to the corresponding hour of the next day, either for a single day or for the mean of a group of days, is called the *non-cyclic* variation. At many observatories it is the custom to take 25 hourly means, for 0^h to 24^h, in forming mean daily inequalities, and to modify the sequence by removing the non-cyclic variation calculated as if it proceeded at a uniform rate during the 24-hourly interval; the values at 0^h and 24^h are thus made equal, and the sequence may then be given equally well either for 0^h to 23^h or 1^h to 24^h (see 16.2).

Schmidt [78] introduced in the Potsdam year-book an additional tabulation (and a diagram) based on 24-hourly means of the components X, Y, and Z for daily intervals beginning at 0, 6, 12, and 18 hours Greenwich time. These daily means are expressed as 'deviations from the normal value' by subtracting the annual mean for the year centred at the particular day. Graphs of this kind are also given in the year-books of other observatories. They provide a smoothed version of the original records (see Fig. 5 on p. 364).

2.15. The harmonic expression of daily inequalities. Some of the principal magnetic observatories harmonically analyse the daily inequalities, often with the non-cyclic variation eliminated (which is desirable). The Fourier coefficients are usually given, in publications on geomagnetism, for frequencies $n = 1, 2, 3, 4$, relative to the formulae

$$\sum (a_n \cos nt + b_n \sin nt) \tag{17}$$

or

$$\sum c_n \sin(nt + \epsilon_n), \tag{18}$$

where t denotes local mean time reckoned in angular measure at the rate of $15°$ per hour, local mean midnight being the initial epoch. Fourier coefficients of solar daily inequalities will in this book always refer to formulae of the types (17) or (18) (see Chapter XVI).

If it is desired to express the Fourier series (18) relative to local *apparent* solar time t' (§ 13), the equation of time e must be allowed for. The formula becomes $\sum c_n \sin(nt' + \epsilon'_n)$, where (§ 13)

$$\epsilon'_n = \epsilon_n + ne, \qquad (19)$$

and e has the following values for the mean epoch in each of the twelve months of the year:

Jan. $+2°$ $19'$	April $+0°$ $4'$	July $+1°$ $21'$	Oct. $-3°$ $28'$
Feb. $+3°$ $29'$	May $-0°$ $52'$	Aug. $+0°$ $59'$	Nov. $-3°$ $47'$
Mar. $+2°$ $12'$	June $+0°$ $4'$	Sept. $-1°$ $11'$	Dec. $-1°$ $6'$

2.16. Variables of position used in the description of the earth's magnetic field. The earth's magnetic field near the surface appears to revolve with the earth: the inequalities in its distribution round any parallel of latitude remain constant at each longitude, apart from the secular variation. Hence the distribution of the main field over the surface is best specified in terms of the position-variables λ (the longitude measured positively to the east from Greenwich) and l (the latitude measured positively to the north) or θ (the north polar distance $\frac{1}{2}\pi - l$).

The variation-field $\Delta_i \mathsf{F}$ (1.6) during any interval is a function of position and time. Its distribution over the surface can be specified in terms of λ, θ, and the Greenwich time t_0, but for an interval such as a month, in which the secular variation is small relative to the transient variations (6.1), an alternative set of variables is preferable: this set is λ', θ, t_0, where λ' denotes the local apparent time t' regarded as a position-variable—that is, as the longitude of the point P at the time t_0, relative to the meridian (C'N in Fig. 16) opposite to that containing the sun. This meridian is in motion relative to the earth. When the local apparent time t' (§ 13) is thus used as a measure of longitude (whether expressed in degrees or in hours) the symbol λ' will be used instead of t: $\lambda' = t' = \lambda + t'_0 = \lambda + t_0 - e$. The difference between apparent and mean local time may often be neglected.

No theoretical assumption is implied by the choice of λ' instead of λ; either variable is simply related to the other, and the distribution expressed in terms of λ', θ, and t_0 may readily be expressed in terms of λ, θ, and t_0. The choice of λ' is often preferable, however, because the intensity of the variation-field $\Delta_i \mathsf{F}$ at any point P is found to depend

closely upon the situation of P relative to the sun, and to show similar features at different stations in the same latitude, at the different times at which in turn they occupy the same position relative to the sun.

The instantaneous world-wide distribution of a magnetic field (or other function of position) specified in terms of λ', θ, and t_0 represents the distribution as it would appear to a distant observer on the sun. It is convenient to suppose him to view the earth with its north pole uppermost. The visible hemisphere would be sunlit, and the invisible one dark. The 'great circle' bounding the visible disk, and separating the sunlit from the dark hemisphere, may be called the *twilight circle*. Stations crossing it, owing to the earth's daily rotation, pass from light to darkness or vice versa. Only one pole of rotation, N or S, will be visible except at the equinoxes, when both will be on the edge of the disk. The position of the poles on the disk will change slowly owing to the earth's orbital motion (see Fig. 1 on p. 160).

The variables λ', θ specify the position of terrestrial points most simply, relative to such an observer, at the equinoxes. The twilight circle then consists of two meridians: for stations on the left-hand one $\lambda' = 6^h$, and on the right hand $\lambda' = 18^h$. The central meridian corresponds to 12^h (noon), and the opposite one (partly visible except at the equinoxes) corresponds to 0^h or 24^h (midnight). Over the left-hand half of the earth λ' varies from 0^h to 12^h, and over the right from 12^h to 24^h: these hemispheres may be called the a.m. and p.m. (*ante or post meridiem*) hemispheres. The four quadrants of the earth between 0^h and 6^h, 6^h and 12^h, 12^h and 18^h, 18^h and 24^h may be numbered in order, 1 to 4. The local time λ' indicates in which quadrant a station lies at any instant, and together with the north polar distance θ' it specifies its exact position in the quadrant.

At non-equinoctial epochs the twilight circle is not composed of meridians, and the four quadrants are each partly sunlit and partly dark.

2.17. International co-operation in geomagnetism, and the International Polar Years. The present *International Meteorological Organization*, in various forms, has existed since 1872; among its former names are 'International Meteorological Committee' or 'Conference of Directors' (see the brief historical account by Hellmann [G 60]). At the International Meteorological Conference at Paris in 1896 this organization constituted a permanent 'Commission for Terrestrial Magnetism and Atmospheric Electricity', which has held conferences as follows: Paris 1896, Bristol 1898, Paris 1900, Innsbruck 1905, Berlin

1910, Utrecht 1923, Zürich 1926, Copenhagen 1929, Innsbruck 1931, Warsaw 1935. The International Meteorological Organization has had a permanent secretariat since 1928, supported by about forty contributing nations; it publishes the reports of the meetings and also the International Magnetic Character Figures (prepared by G. van Dijk at the Kon. Nederlandsch Meteorologisch Institut, De Bilt, near Utrecht). The discussions and, especially, the resolutions adopted by the Commission are of great value to magnetic observatories. A *Kodex* [G 60] collecting the resolutions up to 1910 has been published; and a new edition is being prepared.

A separate body, the *International Geodetic and Geophysical Union*, was founded in 1919. One of its seven sections is the Association of Terrestrial Magnetism and Electricity. Meetings of the Union and its Associations have been held as follows: Brussels 1919, Rome 1922, Madrid 1924, Prague 1927, Stockholm 1930, Lisbon 1933, Edinburgh 1936, Washington 1939. The Association publishes Bulletins (see refs. [66–72] in the General Bibliography) containing the Transactions of the meetings, including resolutions, national reports, special reports of the various committees, and communications.

Among its present committees are the following: On the selection of sites for new observatories for terrestrial magnetism and electricity; On the Aurora; On the study of the relationship between solar activity and terrestrial magnetism; On magnetic secular-variation stations; On magnetic charts; On registration in Iceland of giant pulsations; On methods of observatory publication; On classification of magnetic literature; On the international comparison of magnetic standards; On methods and codes to describe magnetic disturbances and perturbations; and (jointly with the International Scientific Radio Union) On studies of the ionosphere.

The International Council of Scientific Unions (formerly the International Research Council) appointed, in 1924, a *Commission to further the study of Solar and Terrestrial Relationships*: this Commission has published five reports during the years 1926–39 [G 86–9].

These various organizations materially assist international collaboration in observation and research in geophysics. Their publications are widely distributed, to national and university libraries, and to the libraries of magnetic observatories and other scientific institutions. The place of publication and the stock of copies of their works may change from time to time, according to the country or institution with which the secretary for the time being is connected. Where there is difficulty

in consulting or purchasing these international publications, help or advice may be available on application to libraries or magnetic observatories.

An *International Polar Commission* was appointed in 1879 by the International Meteorological Committee in order to organize the occupation of temporary meteorological and magnetical observatories in the north polar region for one year, 1882–3, in accordance with a plan proposed by Weyprecht. Many nations responded, and the results of the observations at these polar stations, as well as the data obtained at permanent observatories on the previously agreed *international days*, have been published in many volumes.

A similar co-operation for the south polar area, on a smaller scale, was carried out by several Antarctic expeditions 1902–3, supplemented by Birkeland's stations in the north.

The *Second International Polar Year* 1932–3 was organized fifty years after the first by the International Polar Year Commission (President: D. La Cour). During thirteen months (August 1932 to August 1933) an extensive programme was carried out at many temporary and other observatories all over the globe. North of 55° latitude, sixteen permanent and twenty-four temporary magnetic stations were in operation. The Central Bureau of the Commission, connected with the Danish Meteorological Office, has collected copies of nearly all the magnetic and earth-current records made during the Polar Year. These can be made available to investigators elsewhere. A file of the records, copied, in highly reduced form, on ordinary motion-picture film, makes it possible to condense an enormous amount of information in a very small space.

2.18. The Department of Terrestrial Magnetism (Carnegie Institution of Washington). In 1904 the Carnegie Institution of Washington founded a Department of Terrestrial Magnetism and appointed Dr. L. A. Bauer, who had initiated the proposal for the Department, as first Director; he was later succeeded by Dr. J. A. Fleming. This Department has its central office at 5241 Broad Branch Road, Washington, D.C., U.S.A.

Its first great project for the advancement of geomagnetism was to complete a comprehensive world magnetic survey by filling up the large gaps then existing in magnetic data by means of land and (still more) ocean expeditions. Its first ocean surveys were made with a wooden sailing vessel, the *Galilee*; then a non-magnetic vessel, the *Carnegie*, was specially constructed, which made extensive surveys over many years; this vessel was unfortunately lost by fire, with its commander

Ault, at Apia, Samoa, in 1929. The Department has undertaken or collaborated in numerous land magnetic surveys; and it has instituted and maintains two magnetic observatories, at Watheroo (Western Australia) and at Huancayo (Peru), where also many other associated kinds of geophysical investigation are carried on. The Department supplements its magnetic work by direct radio-exploration of the *ionosphere*; the pioneer 'pulse' reflection experiments of Breit and Tuve in 1926, made under the auspices of the Department, are now carried on extensively by the automatic multifrequency method developed by Berkner (Ch. XV).

By observation and theoretical research, and by widespread encouragement, advice, and help to geomagnetic undertakings the world over, the Department has been and is of unique service to the science. Some few particulars of its extensive publications are given in the bibliography [G 44].

2.19. Stereoscopic drawings. All three components of the magnetic field-vector vary in time and in space. Ordinary plane drawings and curves are inadequate to show the variations of F, and much training and imaginative power are needed in order to gain from the published data a true conception of the vectorial changes. Wire models have been used to represent the changes, and stereoscopic drawings, which can be easily reproduced and manifolded, have also proved satisfactory for the purpose. Formulae and practical rules for constructing stereograms and anaglyphs (stereoscopic drawings in two colours) have been given by Bartels [79], who has applied the method to ten examples, which convey a vivid impression of such phenomena as the daily and annual magnetic variations, the differences of the mean vector on disturbed and quiet days, a world chart of secular variation, and a globe showing the magnetic axes of the earth and the distribution of magnetic observatories. Peters has constructed stereograms for the field vectors at the commencement of a magnetic storm [9.12]. Wider use of such stereograms seems desirable.

Fig. 17 [79] illustrates the globe in this way; it indicates the intersections of latitude-circles 0°, 30°, and 60° with meridians 0°, 15°, 30°,.... The Greenwich meridian is nearest to the observer. Forty-five observatories are indicated by black circles and initials. The axes are: NS = the earth's axis of rotation; AB = the earth's diameter parallel to the magnetic axis of the magnetic dipole which approximates most closely to the earth's field; PP = the line joining the two poles of magnetic dip; this line passes the earth's centre at a distance of 1,140 km.

The diagram shows the complete lack of observatories in the African

sector south of 30° north between 45° west and 45° east (nearly one-fifth of the surface of the earth) at the time when Fig. 17 was drawn (since then a magnetic observatory has been reinstituted at Capetown). It also illustrates the importance of such isolated stations as Apia (Samoa) (marked A) and Honolulu (H). The network of stations is closest in
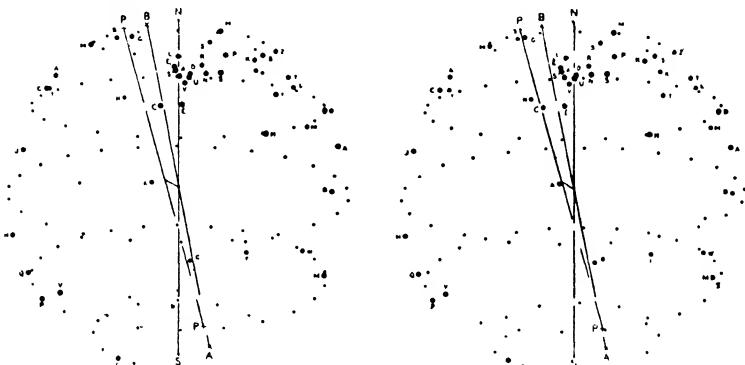
FIG. 17. Stereogram showing the earth's rotational and magnetic axes, and the distribution of the magnetic observatories

Europe. The American sector contains a fairly good chain of observatories.

2.20. Special pantographs. At different observatories the scale-values of the magnetographs differ, even for the same element, and they may be unlike for different elements at the same observatory. Sometimes even reversed curves are recorded, with increase of H and Z indicated by opposite slopes of the curves. This adds to the inherent difficulty of comparing the field changes at different observatories, for example in the study of a single magnetic storm. Ordinary photographic enlargement, or reduction in size, is an ineffective remedy, because in order to change the curves to the internationally adopted standards (1'/mm. in D, 5γ/mm. in the force components, 15 mm./hour in time) the scales of abscissae and ordinates must be changed in different ratios.

Schmidt constructed a satisfactory special pantograph meeting these requirements; the Potsdam curves are reproduced by it as a matter of routine, and all the magnetograms of the German Antarctic Expedition [G 96] were reproduced by it and published in an atlas. Unfortunately no other observatory has followed this practice. The pantograph, which is hand-operated, has been described by K. Luyken [81]. Fig. 18, part of a more extensive example given in the Potsdam year-book for 1922, shows not only the possibilities afforded by Schmidt's pantograph but also, by the different aspect of the curves, the need for its application.

Peters and Green [83] invented an ingenious simple photographic method with the same object. The magnetogram, with a rectangular grid placed over it, is photographed by a camera in such a position that the (horizontal) optical axis of the lens is perpendicular to the horizontal axis OX of abscissae. The grid can be inclined backwards around the axis OX. The resultant negative shows a perspective image of the
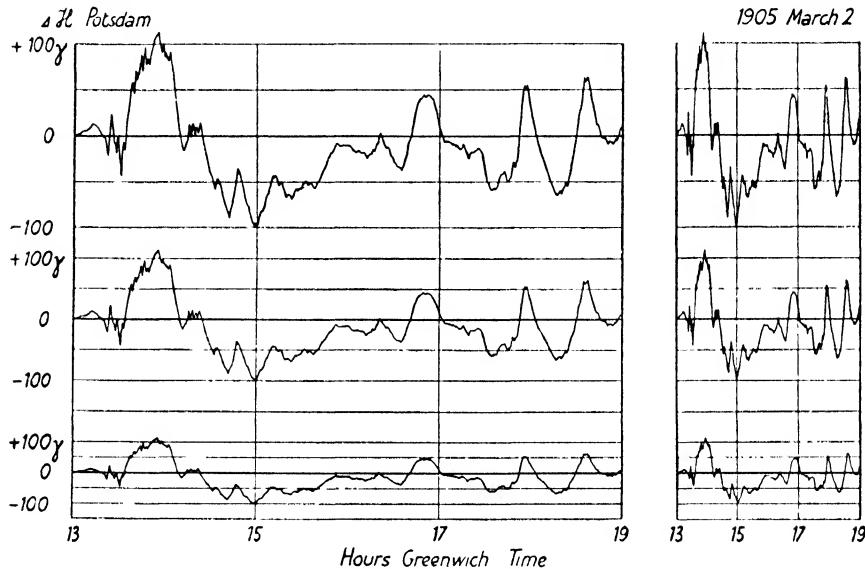
FIG. 18. Variations ΔH of the horizontal intensity, recorded at Potsdam, 1905 March 2, reproduced in six diagrams having different ratios of the scales for H (ordinates) and time (abscissae); drawn with the aid of Schmidt's pantograph

grid; lines which in the original were parallel to the ordinate axis OY, converge in the photograph. This negative is again photographed, with suitable forward inclination around the same axis, so that the lines parallel to OY become parallel again. The result of the two exposures (made with suitably small stops to obtain sharp images) is a rectangular grid with a reduced ratio of ordinates to abscissae. With this procedure irrelevant lines on the magnetogram may be obliterated on the first negative. An example of this method is reproduced as Fig. 37 on p. 336.

B. SOME DETAILS OF MAGNETIC MEASUREMENTS

(This part uses mathematical developments described in Chapters XVI, XVII, and XVIII.)

2.21. The parameters of a magnet. The field of a magnet, in the non-magnetic space round it where its magnetic potential V satisfies Laplace's equation $\nabla^2 V = 0$ (1.7), may be specified by the expression

of V as a series of spherical harmonics, relative to a polar coordinate system, with some origin O, in terms of the radius r, the polar angular distance σ, and the longitude τ. Using normalized spherical harmonics, as described in Chapter XVII, we may therefore write (corresponding to 17 (17) and 18 (1, 2), but not referring to the outside of any particular sphere of radius $r = a$; also here $V^e = 0$)

$$V = \sum_{n=1}^{\infty} \sum_{m=0}^{n} (g_n^m \cos m\tau + h_n^m \sin m\tau) P_n^m (\cos \sigma)/r^{n+1} = \sum_n \sum_m V_n^m = \sum V_n.$$
$$(21)$$

We also write $\quad g_n^m \cos m\tau + h_n^m \sin m\tau = c_n^m \sin(m\tau + \gamma_n^m).$ $\qquad(22)$

The coefficients g_n^m, h_n^m have the physical dimensions gauss cm.$^{(n+2)}$ and are called (by Schmidt) the *parameters* of the magnet with respect to the coordinate system chosen. For bar-magnets of the hollow cylindrical shape ordinarily used in magnetic measurements, the origin is chosen as the geometric centre, the axis $\sigma = 0$ as the axis of the cylinder, and $\tau = 0$ as the direction towards a mark engraved on the surface, on the circle $\sigma = 90°$. (The geometric centre in general differs from the *magnetic* centre, relative to which, as origin, the expression for V takes a form (the *canonical* form) simplified by the absence of certain terms. The magnetic centre is not of great importance in magnetic measurements, and its discussion is therefore deferred to 18.6.)

The potential of a single magnet, and the mutual attraction of two magnets, have been most comprehensively treated by Schmidt [84–92], and the next sections give an outline of his theory, which is also of importance in improving the standard of absolute observations of H.

2.22. The regular magnet. Due to inhomogeneities in the steel and other circumstances, the potential V will, in general, not be of a simple form, and the symmetry of the geometrical shape of a magnet is not preserved in V. A simplification may be reached in the following way: Consider the ν meridian planes through the geometrical axis, at equal azimuthal angles $\rho = 2\pi/\nu$, where ν is any integer; let $V(1)$, $V(2),..., V(\nu)$ denote the values of V in these meridian planes $\tau = \rho$, $\tau = 2\rho,..., \tau = \nu\rho$. The expressions for these values of V may be written as follows:

$$V(1) = \sum_{m=0}^{\infty} (G_m \cos m\rho + H_m \sin m\rho), \quad ...,$$

$$V(\nu) = \sum_{m=0}^{\infty} (G_m \cos m\nu\rho + H_m \sin m\nu\rho), \quad (23)$$

where G_m and H_m depend on r and σ. The ordinary harmonic analysis of $V(1),..., V(\nu)$ would determine G_m and H_m up to $m \leqslant \frac{1}{2}\nu$ (16.12 and

14), and this is sufficient if the terms for all higher values of m are negligible. The actual values of G_m and H_m could be found (for any chosen values of r and σ) by analysis of measurements of V in successive positions of the magnet, turned successively around its axis of symmetry, into the positions $\tau = -\rho$, $\tau = -2\rho, ..., \tau = -\nu\rho$, V then being measured in the plane corresponding to $\tau = 0$. This would yield $V(1), ..., V(\nu)$. These ν positions are called collimation-positions by Schmidt.

From 16 (45 a), with the reservations mentioned in 16.14 concerning the influence of the higher terms in m, it follows that

$$G_0 = [V(1) + V(2) + ... + V(\nu)]/\nu. \tag{24}$$

In the average of the ν positions, the magnet has therefore the same field as a magnet of axial symmetry, for which

$$V = \sum_{n=1}^{\infty} g_n^0 P_n^0(\cos \sigma)/r^{n+1}. \tag{25}$$

In the ordinary observatory practice of H-measurements, ν is taken as 2 (the magnet being placed in the positions with the mark up and the mark down), because it is found that an increased number of positions does not lead to materially greater accuracy.

In addition to these operations, which aim at eliminating the effect of axial asymmetries of the magnet, the magnet is turned around the axis $\tau = 0$ by an angle π, so that the position of the poles is reversed. If V in two such positions, direct and reversed, is denoted by V_+ and V_-, then the expression for V_- can be derived from that for V_+ by changing σ to $(\pi - \sigma)$, $\cos \sigma$ to $-\cos \sigma$; since

$$P_n^0(\cos \sigma) = (-1)^n P_n^0(-\cos \sigma), \tag{26}$$

we have for the half-difference between V_+ and V_-

$$V = \tfrac{1}{2}(V_+ - V_-) = g_1^0 P_1^0(\cos \sigma)/r^2 + g_3^0 P_3^0(\cos \sigma)/r^4 + g_5^0 P_5^0(\cos \sigma)/r^6 + ... \tag{27}$$

The first term corresponds to an equivalent dipole of moment g_1^0 situated at the geometric centre of the magnet (1 (15)). This term will usually be much larger than the remaining terms, which may be regarded as 'corrections' to the first term. If a magnet has a potential of the form (27), it is called *regular*; (27) shows that the *difference* between the potentials of *any* magnet of axial symmetry, in the direct and reversed positions of the poles, is of the same form as if the magnet were regular. If, therefore, only these differences of the potentials are used in the observations, and the magnet is revolved around its axis as already described, in order to eliminate, in the average, any axial

asymmetry, every magnet acts like a regular magnet, and its geometric centre coincides with its magnetic centre.

This result is important, because the magnet can easily be placed accurately in the various positions by suitable stops on the theodolite. In the practice of deflexion experiments for measuring H, the four positions of the deflecting magnet are denoted by 'mark up' or 'down', and 'north pole east' or 'west'. The number of positions is doubled again by placing the deflecting magnet successively east and west of the deflected needle.

2.23. The schematic magnet. Older studies of the field of magnets start with a supposed distribution of magnetization inside the magnet, or, simply, from a distribution of magnetic matter. A much-used generalization of a dipole is the *schematic magnet* [90], consisting of two equal and opposite poles P_+, P_- of strength $\pm\mu$ at a finite distance $2d$ apart. If C is the point midway between them, then the distances of P_+ and P_- from any point P, at a distance r from C, are given by

$$P_+P^2 = d^2 - 2rd\cos\sigma + r^2; \qquad P_-P^2 = d^2 + 2rd\cos\sigma + r^2,$$

where σ denotes the angle PCP_+. Hence the potential at P, by 1 (4) and 17 (32 b), is

$$V = \frac{\mu}{P_+P} - \frac{\mu}{P_-P} = \mu\sum_{n=0}^{\infty}P_n^0(\cos\sigma)d^n/r^{n+1} - \mu\sum_{n=0}^{\infty}P_n^0(-\cos\sigma)d^n/r^{n+1}.$$

Let
$$2\mu d = M, \tag{28}$$

then by (26) the expression for V reduces to

$$V = MP_1^0(\cos\sigma)/r^2 + Md^2P_3^0(\cos\sigma)/r^4 + Md^4P_5^0(\cos\sigma)/r^6 + \dots. \tag{29}$$

Comparison with (27) gives

$$g_1^0 = M, \qquad g_3^0 = Md^2, \qquad g_5^0 = Md^4,\dots, \tag{30}$$

showing that, so far as concerns the first two terms in (27), a regular magnet is equivalent to a schematic magnet of moment M and length

$$2d = 2(g_3^0/M)^{\frac{1}{2}}. \tag{31}$$

Many experiments have been made to determine the ratio of $2d$ to the geometrical length l of the magnet; for ordinary cylindrical magnets, according to Börgen [92] and Kohlrausch [92 a],

$$0{\cdot}80l < 2d < 0{\cdot}88l. \tag{31 a}$$

By comparing the formulae (27) and (29), it becomes clear that the potential of a regular magnet up to the first $2n$ (or $(2n+1)$) terms can be interpreted as a superposition of n schematic magnets (or n schematic magnets and one dipole) arranged along the same axis. Schmidt [90] has discussed this interpretation.

2.24. Equivalent magnetic distributions. It may be shown that, outside a sphere of radius a which encloses a magnet completely, the potential V of the magnet, as given by equation (21), is identical with that of a surface-distribution of hypothetical (positive or negative) magnetic matter on this sphere, of density

$$f(\sigma,\tau) = \sum_{n=1}^{\infty} \sum_{m=0}^{n} (G_n^m \cos m\tau + H_n^m \sin m\tau) P_n^m(\cos\sigma); \qquad (32)$$

$f(\sigma,\tau)$ is called the *equivalent magnetic distribution* over the sphere a. This is a direct consequence of the theorem proved in 17.16; from a comparison of formulae (71) and (74) given there with our equation (21), it follows that it is only necessary to put

$$G_n^m a^{n+2} 4\pi/(2n+1) = g_n^m, \qquad H_n^m a^{n+2} 4\pi/(2n+1) = h_n^m. \qquad (33)$$

2.25. The mutual potential energy of two dipoles. In the case of dipoles the expression for the mutual potential W can be derived directly from first principles as follows. Let M_1, M_2 denote their (vector) moments, and O_1, O_2 their centres. Let d denote the vector $O_1 O_2$, of length d equal to the mutual distance $O_1 O_2$. Let the co-ordinates of any point P referred to O be r, σ_1, τ_1, the angular coordinates being referred to the direction of M_1 as polar axis. Then the potential V_1 of M_1 at P is $M_1 \cos\sigma_1/r_1^2$. Let the dipole M_2 be regarded as a system of two magnetic poles, m_2 at O_2' and $-m_2$ at O_2'', where $O_2'' O_2$ and $O_2 O_2'$ are in the direction of M_2, and each is of length b_2; we suppose that $2m_2 b_2 = M_2$, and in the limit take b_2 infinitesimal. Then the mutual potential energy, defined as the energy W required to bring the dipole M_2 to its given position in the field of M_1, is the work needed to bring m_2 and $-m_2$ into their positions O_2' and O_2'', against the field of M_1; that is,

$$W = m_2 V_1' + (-m_2) V_1'' = m_2(V_1' - V_1''), \qquad (34)$$

where V_1' and V_1'' are the values of V_1 at O_2' and O_2''. Now $V_1' - V_1''$ is the difference of potential (due to M_1) between O_2' and O_2'', and $(V_1' - V_1'')/2b_2$ is the component along $O_2'' O_2$ of the magnetic intensity at O_2 due to M_1. If E_1 is this intensity, and θ is the angle it makes with M_2 or $O_2'' O_2$, then

$$\frac{V_1' - V_1''}{2b_2} = E_1 \cos\theta. \qquad (35)$$

Hence

$$W = m_2(V_1' - V_1'') = 2m_2 b_2 \frac{V_1' - V_1''}{2b_2} = M_2 E_1 \cos\theta = M_2 \cdot E_1, \qquad (36)$$

using (in the last expression) the notation for a scalar product of the vectors.

At any point P, whose position relative to O_1 is represented by the vector r_1, E_1 can be expressed in the following form:

$$E_1 = -M_1/r_1^3 + 3(M_1 \cdot r_1)r_1/r_1^5; \qquad (37)$$

it is easily verified that this is identical with the usual formulae 1 (16), which are the component equations of (37) in the directions of R and S there. Formula (37) implies that E_1 can be represented as the sum of two components, one, of magnitude M_1/r_1^3, in the direction opposite to that of M_1 itself, and the other, of magnitude $(3M_1 \cos \sigma_1)/r_1^3$, along $O_1 P$.

Substituting this expression for E_1 in (36), and remembering that at O_2, $r_1 = d$, we have

$$W = \frac{-(M_1 \cdot M_2)}{d^3} + \frac{3(M_1 \cdot d)(M_2 \cdot d)}{d^5} = \frac{M_1 M_2}{d^3}(3 \cos \sigma_1 \cos \sigma_2 - \cos \epsilon), \quad (38)$$

where σ_1, σ_2 are the polar angles of $O_1 O_2$ (not of $O_2 O_1$) relative to the directions of M_1 and M_2 respectively, and ϵ is the angle between the directions of M_1 and M_2.

Suppose that instead of taking special directions of the polar axis for coordinates relative to O_1 and O_2 we take the polar axis in both cases along the z-axis of an orthogonal coordinate system Ox, Oy, Oz, and measure azimuths from the plane xOz. Then if θ_1, ϕ_1, and θ_2, ϕ_2 are the angular coordinates of the moments M_1 and M_2, and if θ, ϕ are those of $O_1 O_2$, it is readily seen that

$$\cos \epsilon = \cos \theta_1 \cos \theta_2 + \sin \theta_1 \sin \theta_2 \cos(\phi_1 - \phi_2), \qquad (39)$$

$$\cos \sigma_1 = \cos \theta_1 \cos \theta + \sin \theta_1 \sin \theta \cos(\phi_1 - \phi), \qquad (40)$$

$$\cos \sigma_2 = \cos \theta_2 \cos \theta + \sin \theta_2 \sin \theta \cos(\phi_2 - \phi). \qquad (41)$$

Moreover $\qquad d^2 = (x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2. \qquad (42)$

2.26. The mutual force between two dipoles or magnets. Let X_1, Y_1, Z_1 denote the components of the mechanical force, and L_1, M_1, N_1 those of the mechanical couple (about O_1) exerted by M_2 on M_1. If M_1 is moved slightly, without change of direction, so that its coordinates are changed from x_1, y_1, z_1 to $x_1 + \delta x_1, y_1 + \delta y_1, z_1 + \delta z_1$, a force $-X_1$, $-Y_1$, $-Z_1$ must be applied to it to balance the force due to M_2. The work done on M_1 is therefore

$$-(X_1 \delta x_1 + Y_1 \delta y_1 + Z_1 \delta z_1),$$

and the mutual potential energy W will be increased by this amount, which we may denote by δW. Hence

$$\delta W = -(X_1 \delta x_1 + Y_1 \delta y_1 + Z_1 \delta z_1),$$

whence it follows that

$$X_1 = -\frac{\partial W}{\partial x_1}, \qquad Y_1 = -\frac{\partial W}{\partial y_1}, \qquad Z_1 = -\frac{\partial W}{\partial z_1}. \tag{43}$$

The force X_2, Y_2, Z_2 upon M_2 due to M_1 is similarly given by

$$X_2 = -\frac{\partial W}{\partial x_2}, \qquad Y_2 = -\frac{\partial W}{\partial y_2}, \qquad Z_2 = -\frac{\partial W}{\partial z_2}.$$

Since W involves the coordinates x_1, y_1, z_1 and x_2, y_2, z_2 only through its dependence upon the distance $O_1 O_2$, it is easy to see (from (42)) that (for example)

$$\frac{\partial W}{\partial x_1} = -\frac{\partial W}{\partial x_2} \quad \text{or} \quad X_1 = -X_2, \tag{44}$$

that is, the forces X_1, Y_1, Z_1 and X_2, Y_2, Z_2 are equal and opposite.

These remarks apply also to any two general magnets, such as are discussed in §§ 21, 24, and 28.

In the case of the two dipoles, it is not difficult to deduce the following vector expression for X_1, Y_1, Z_1:

$$\frac{3M_1 M_2}{d^5}(\cos\epsilon - 5\cos\sigma_1\cos\sigma_2)\mathbf{d} - \frac{3}{d^4}(M_1\cos\sigma_1\,\mathbf{M_2} + M_2\cos\sigma_2\,\mathbf{M_1}). \tag{45}$$

This expression gives the force as the sum of three parts, in the directions of $\mathbf{d}$, $\mathbf{M_1}$, and $\mathbf{M_2}$.

2.27. The couple exerted by one dipole or magnet upon another. If in the expression (38) for W, the mutual potential energy of two dipoles, we substitute from (39)–(41), W is obtained as a function of $x_1, y_1, z_1, x_2, y_2, z_2, \theta_1, \phi_1, \theta_2, \phi_2, \theta, \phi$. If we imagine that ϕ_1 is changed slightly, to $\phi_1 + \delta\phi_1$, all the other variables remaining constant, W will change by a small amount δW such that

$$\delta W = \frac{\partial W}{\partial \phi_1}\delta\phi_1. \tag{46}$$

Such a change of ϕ_1 corresponds to a rotation of the dipole M_1 about $O_1 z$ (an axis through O_1, parallel to Oz), through the angle $\delta\phi_1$. To produce such a rotation, a couple $-N_1$ about $O_1 z$ must be applied to the dipole M_1 to balance the z-component (N_1) of the couple exerted on M_1 by M_2. Thus the work done on M_1, increasing W by δW, is $-N_1\delta\phi_1$; consequently

$$\delta W = -N_1\delta\phi_1,$$

and therefore

$$N_1 = -\frac{\partial W}{\partial \phi_1}. \tag{47}$$

Similarly $N_2 = -\partial W/\partial\phi_2$.

The other components of the couples, about axes parallel to Ox and Oy, may be obtained by expressing W in terms of angular coordinates relative to Ox or Oy as polar axes.

Similar remarks apply to the couple components acting on general magnets, when the expression for W is expressed in terms of coordinates which include those whose variation corresponds to a simple rotation of one or other magnet about one of the coordinate axes.

In the case of the two dipoles it is possible, more directly, by vector methods, to express the couple L_1, M_1, N_1 on M_1 in the following vector form (involving vector products—cf. 1.10):

$$3(M_2 \cdot d)M_1 \wedge d/d^5 - M_1 \wedge M_2/d^3. \tag{48}$$

The components of this in any direction can be easily written down.

2.28. The interaction of two magnets, general expression. Consider the general case of two magnets 1 and 2, each of which has a potential (V_1 and V_2) of the general form (21). It may be possible to enclose the magnets in spheres which do not intersect; the equivalent magnetic distributions (§ 24) may be f_1 and f_2, and dS_1 and dS_2 may denote the surface elements of the spheres, so that $f_1 dS_1$ is the pole-strength in the area dS_1. Now the mutual potential energy between a magnet with the potential V_1 and a magnetic pole of unit strength (positive or negative) is V_1: between the magnet and a pole of strength m it is $V_1 m$. Because of the identity of the potentials V_1 and V_2 of the magnets with the potentials of the distributions f_1 and f_2, the mutual potential energy W is given by the two alternative expressions

$$W = \int V_1 f_2 \, dS_2 = \int V_2 f_1 \, dS_1, \tag{49}$$

where the integrations are extended over the surfaces of the spheres 2 or 1.

In evaluating the first integral the angular coordinates used would be σ_2, τ_2 relative to the centre O_2 (with coordinates x_2, y_2, z_2) of the sphere S_2; f_2 may be supposed known in terms of these coordinates, but in the first instance V_1 will be expressed in terms of the coordinates r_1, σ_1, τ_1 relative to the centre O_1 (at x_1, y_1, z_1) of S_1. Hence before integration V_1 must be transformed so as to refer to the coordinates (a_2, σ_2, τ_2) of any point on the sphere S_2 (of radius a_2). This involves general transformation formulae for spherical harmonic functions, for change of origin, and (possibly) of the directions of reference for angular coordinates. Such general formulae have been worked out by Schmidt [84–90] and applied to this problem.

2.29. Deflecting bar and needle in selected positions. In magnetic measurements the field of a deflexion magnet (referred to as the bar) or of a coil M_1, situated at O_1, is measured by its action on a magnetic needle M_2, situated at O_2. Three principal relative positions of the line $O_1 O_2$ connecting the centres of the bar and the needle, and of their magnetic axes, are used. These can be described, with respect to Fig. 19, as follows:

First position: M_1, directed along $O_1 x_1$, acts on M_2, directed along
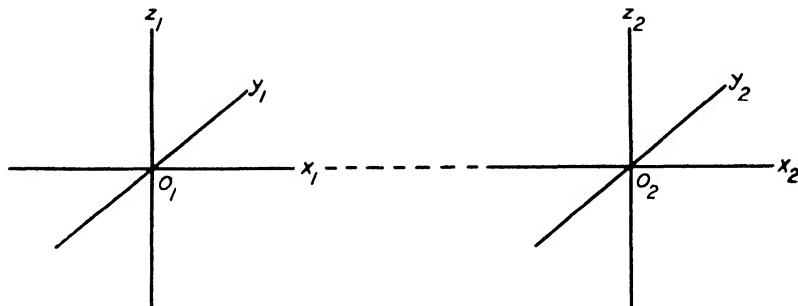
FIG. 19. Coordinate systems for the bar 1 and the needle 2

$O_2 y_2$; $O_1 O_2$ lies along $O_1 x_1$; the needle turns around the axis $O_2 z_2$. This is Lamont's position, as used in measuring H by the magnetic theodolite.

Second position: M_1, directed along $O_1 y_1$, acts on M_2, directed along $O_2 y_2$; the needle turns around the axis $O_2 z_2$.

Third position: M_1, directed along $O_1 z_1$, acts on M_2, directed along $O_2 y_2$; the needle turns around the axis $O_2 x_2$.

We shall give here the mechanical couples (denoted by A, B, C) for these three positions only. The direction of the couples may be seen from the fact that from the first position stable equilibrium is reached if the needle is turned parallel to the bar, and from the two other positions, if it is turned anti-parallel to the bar.

Suppose that both the bar and the needle are regular magnets, whose parameters, in the standard form (27) (but omitting the upper index 0), are g_1, g_3, g_5,... for the bar and j_1, j_3, j_5,... for the needle; let e be the distance between their centres. Then

$$A = 2g_1 j_1/e^3 + (4g_3 j_1 - 6g_1 j_3)/e^5 + (6g_5 j_1 - 30g_3 j_3 + \tfrac{45}{4} g_1 j_5)/e^7 + ..., \quad (50)$$

$$B = g_1 j_1/e^3 - (\tfrac{3}{2} g_3 j_1 - 6g_1 j_3)/e^5 + (\tfrac{15}{8} g_5 j_1 - \tfrac{45}{2} g_3 j_3 + 15 g_1 j_5)/e^7 + ..., \quad (51)$$

$$C = g_1 j_1/e^3 - \tfrac{3}{2}(g_3 j_1 + g_1 j_3)/e^5 + \tfrac{15}{8}(g_5 j_1 + 2g_3 j_3 + g_1 j_5)/e^7 - .... \quad (52)$$

PLATE 11
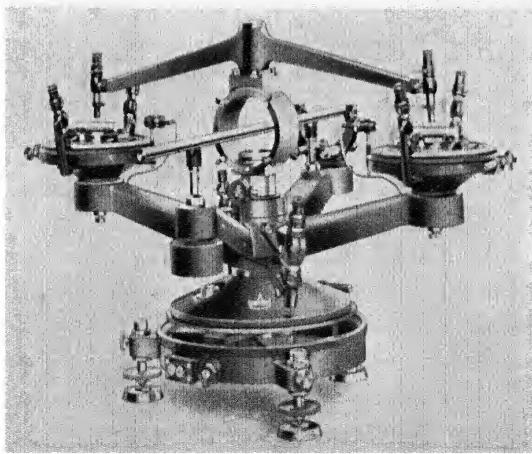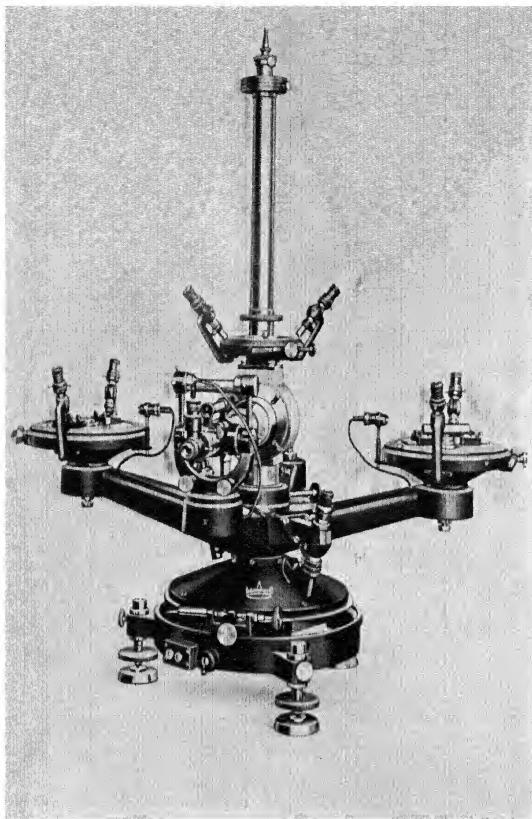
A. Schmidt's theodolite: above, in use for magnetic measurements; below, with its torsion-head replaced by a comparator device for measuring distances (Askania)

If the bar and the needle can be regarded as schematic magnets, M_1 and $2d_1$ being the moment and length of the first, and M_2 and $2d_2$ those of the second, we may insert the parametric values (30) and obtain for Lamont's position

$$A = \frac{2M_1 M_2}{e^3}[1+(2d_1^2-3d_2^2)/e^2+(3d_1^4-15d_1^2 d_2^2+\tfrac{45}{8} d_2^4)/e^4+...]. \quad (53)$$

Schmidt has constructed a theodolite (Plate 11) for the exact determination of the parameters of a deflexion magnet. In this instrument, which has been described by Bock [12, 89], the magnet can take various measurable azimuths relative to the line $O_1 O_2$. Schmidt has also developed the theory of galvanometric determinations of the parameters [88].

2.30. The deflexion constant. In the deflexion experiment with Lamont's theodolite (§ 4), the couple A acting on the needle is balanced by the couple $M_2 H \sin\beta$ due to the horizontal magnetic force. The deflexion constant k in (7) is therefore the bracket in our last formula (53), or the more general corresponding expression obtainable from (50). It is usually written

$$k = 1+\frac{p}{e^2}+\frac{q}{e^4}, \quad (54)$$

where $\qquad p = 2d_1^2 - 3d_2^2, \qquad q = 3d_1^4 - 15d_1^2 d_2^2 + \tfrac{45}{8}d_2^4.$

It is customary to choose the ratio of the lengths d_2/d_1 of the needle and bar so that $q = 0$; this condition is clearly a quadratic equation in $(d_1/d_2)^2$, from which we obtain the solution

$$d_2 = 0\cdot467d_1. \quad (55)$$

Since the ratio (31 a) between the pole-distance $2d$ and the length l is, for magnets of similar shape, nearly constant, we have also $l_2 = 0\cdot467l_1$. For the Potsdam theodolite 'Wanschaff', $l_1 = 7\cdot451$ cm., $l_2 = 3\cdot510$ cm. (so that $l_2 = 0\cdot471l_1$), $e = 29\cdot9982$ cm. The coefficient p, and therefore k, is usually determined (once for all) by measuring β for deflexions at two distances, e and E, for which the best ratio is $E/e = 1\cdot32$. For magnet 2 at Potsdam, $k = 1\cdot01527$ (with standard error $0\cdot00034$); this value of k leads to the ratio $d_1/l_1 = d_2/l_2 = 0\cdot86$ in (31 a).

In the actual deflexion experiment four positions are used: (1) the bar east of the needle, with its north pole east; (2) the bar west, the north pole east; (3) the bar west, the north pole west; (4) the bar east, the north pole west. They yield the deflexion angles $\beta_1, \beta_2, \beta_3, \beta_4$. This completes, with the mark on the bar upwards, a set of deflexion

observations; the oscillations follow, then the bar is turned through 180° about its own axis, so that the mark is downwards, and another oscillation experiment is made, and finally a second set of deflexions. This eliminates terms of the first order in the effect of possible eccentricities of the bar or needle, or a remanent torsion of the suspension of the needle.

A second-order term in the eccentricities is corrected by the formula

$$\beta = \beta' - a\,\Delta\beta^2, \tag{56}$$

where $\beta' = \frac{1}{4}(\beta_1+\beta_2+\beta_3+\beta_4)$, $\Delta\beta = \frac{1}{4}(\beta_1+\beta_2-\beta_3-\beta_4)$ (in arc measure), and $a = \tan\beta'/8\sin^4\frac{1}{2}\beta'$.

The theodolite is fitted with devices which allow the needle to be lifted, or shifted sideways, so that its centre will be accurately on the extension of the axis of the deflexion magnet, and so that the needle can be set perpendicular to the magnet. This condition can be sufficiently well satisfied, because in the mean of the readings for the eight positions, first-order errors are eliminated; the centre of the needle need therefore only be at the correct place within a few thousandths of the distance e, and the angle may differ from 90° by a few tenths of a degree. The correct position (for instance, the correct height of the needle) gives the maximum deflexion, and, as Schmidt has pointed out, it can therefore be found by a few observations in three slightly differing positions (for instance, three heights of the needle) if the deflexion angles are fitted to a parabolic graph.

At observatories it is the custom to read the declination on the variometer records for the times when deflexions are made, in order to reduce the angles of the needle read on the theodolite to an average standard value of the declination.

2.31. Influence of temperature, induction, and ageing. The temperature affects the deflexion experiments in two ways: by the thermal expansion of the brass arms of the theodolite, and by the decrease of the moment M of the deflexion magnet (the bar), according to the approximate formula $M = M_0(1-\alpha t)$.

Since the bar, during the deflexion experiment and during the oscillations, makes different angles with the earth's magnetic force, the additional moment induced in it by the horizontal component H also varies. If M_0 is the moment at right angles to the horizontal force, it is increased to $M_0(1+k'H)$ during the oscillations, and decreased to $M_0(1-k'H\sin\beta)$ during the deflexions. The induction coefficient k' must therefore be determined. This is done either by the usual method

of Weber (in which the magnet is fixed in the plane of a coil, and the current induced in the coil, when this is turned through 180°, is measured by a ballistic galvanometer), or (better) by a device due to Schmidt and Venske [17]: The magnet is placed at the centre C, and along the axis, of the space inside two fixed coaxial solenoids, the inner one being shorter than the outer one; a current is sent in opposite directions through the two coils, whose dimensions are calculated so that although at C there is a magnetic field, which modifies the moment of the magnet, at a point P outside, where a needle is suspended by a vertical fibre so as to measure the field of the magnet, the field of the two coils is zero. Thus the current and the inducing field on the magnet can be varied, and the induction coefficient measured, without being affected by any influence of the field of the coils at the needle.

For the Potsdam magnet no. 2 the temperature coefficient α is 0·000321 and the induction coefficient k' is 0·00613.

The moment of a magnet gradually decreases. This process, called *ageing*, depends on the treatment of the magnet during manufacture (all magnets are individuals); artificial ageing (successive heating to 100° C. and cooling) slows down the ageing during the subsequent life of the magnet. Magnet no. 2 at Potsdam, made of tungsten steel, was taken into use in February 1893, when M was 1,188 Γ cm.^{3}; the moment decreased to the following values (averages for December): 1898, 997; 1907, 916; 1921, 859; 1934, 829. The decrease from 1893 to 1907 approximately agrees with the relation

$$dM/dt = -CMe^{-ct}, \tag{57}$$

but in recent years the decrease has not followed this formula (see *Ergebnisse Potsdam*, 1903/4, p. x). Accidental sudden decreases in the moment are mostly followed by a slight recovery.

When the magnets are not actually being used for measurements they are kept alined between other magnets; when they are taken out of the field of the neighbouring magnets their magnetic moment declines by about 1/2000 of its value, but becomes constant after about an hour [21], in the arrangement used at Potsdam.

2.32. Oscillations. In connexion with the oscillation experiment the moment of inertia I of the bar is measured once for all. This is done by means of a non-magnetic cylindrical bar of accurately known moment of inertia I_1 which fits into the hollow of the magnet; a second oscillation experiment then yields another oscillation period T_1, and I is calculated from the relation $T^2 : T_1^2 = I : (I+I_1)$. W. Watson [106]

proposed the use of 'normals of inertia', that is, non-magnetic bars of which the moment of inertia I_1 is determined with a high degree of accuracy, and which are interchanged between the observatories. The moment of inertia of the suspension must also be determined.

The oscillation period T is usually determined accurately enough by the eye-and-ear method: while the half-second beats of a chronometer are counted, the times of passage of the magnet through its zero-position are estimated by mental interpolation. A scale reflected into a telescope by a mirror attached to the magnet (or simply by its polished end-surface) is used to mark the zero-position of the magnet and its successive passages through this position. About twenty successive passages are observed, then the approximate time of the 100th passage is calculated, and the 100th to 120th passages are observed again. Of course, chronographs may be used instead, or automatic counters. With such an apparatus devised by Fanselau [18, 19], Burger [20] has determined the influence of the apparent increase of the inertia of the magnet by the air oscillating with it.

2.33. Final formulae for the measurement of H. The following causes make small corrections necessary in the reduction of the observations made to determine H, as compared with the simple formulae of § 4 based on the representation of the bar and needle as dipoles:

(i) The finite volume of the distribution of magnetization in the bar and needle; this is taken care of by the deflexion constant k, § 30.

(ii) The influence of temperature on the magnetic moment of the bar-magnet (coefficient α), on its length (coefficient of linear expansion ϵ' for steel), on its moment of inertia (coefficient $2\epsilon'$, because the inertia increases as the square of the linear dimensions), and on the dimensions of the theodolite (coefficient of linear expansion ϵ for brass); it is convenient to express the temperatures t' during the deflexion and t'_1 during the oscillations as deviations from a standard temperature.

(iii) The additional moment induced in the bar-magnet by the geomagnetic field (coefficient k').

(iv) The torsion of the suspension of the bar during the oscillations, which acts in the same sense as H in driving the needle back to rest, thereby shortening the oscillation period T; the torsion ratio, of the mechanical couple Θ due to torsion, and the couple MH due to the horizontal intensity, is $\tau = \Theta/MH$; it is determined by the deflexion δ produced when the torsion-head is turned through 2π, $\tau = \delta/(2\pi-\delta)$.

(v) The gaining or losing rate of the chronometer used in the oscilla-

tion experiment (Δs seconds per day, positive if the chronometer is losing time).

(vi) The finite size of the half-amplitude h of the oscillations, in degrees; and finally

(vii) The variations ΔH and ΔH_1 of H during the times of the deflexion and oscillation experiments ($H = H_0 + \Delta H$, where H_0 is the value for the *base-line* of the variometer). The deflexion angle β is supposed to be calculated from readings on the theodolite, individually freed from the variations of declination, and from the influence of eccentricities (56). Then the following expressions serve to give the values for M_0 at the standard temperature, and the base-line value H_0 of the variometer, in terms of the observed deflexion angle β, the oscillation period T observed by the chronometer, and the instrumental constants (the distance e_0 from the bar to the needle on the theodolite, measured as half the distance between the geometric centres of the bar in the positions east and west of the needle, and the moment of inertia I_0 of the bar magnet, both corrected to the standard temperature):

$$\text{Deflexions:} \quad \frac{M_0}{H_0} = \frac{e_0^3 \sin\beta(1+3\epsilon t')(1+\Delta H/H_0)}{2k(1-\alpha t')(1-k'H\sin\beta)}, \tag{58}$$

$$\text{Oscillations:} \quad M_0 H_0 =$$
$$\frac{\pi^2 I_0(1+2\epsilon' t_1')}{T^2(1+\tau)(1-\alpha t_1')(1+2\Delta s/86400)(1+k'H)(1+\Delta H_1/H_0)(1-0{\cdot}0000381\,h^2)}. \tag{59}$$

(The factor $0{\cdot}0000381$ is $2(\sin 1°)^2/16$; see p. 14.)

Further improvement of the absolute accuracy may be possible in two ways: by a better determination of the deflexion constant k (determining the parameters of the magnet, either by Schmidt's special theodolite or by galvanometric methods, § 30); and by a better correction of the oscillation observations for the increase in inertia due to co-oscillating air masses, as suggested by Fanselau [19].

2.34. Coil systems. In order to produce magnetic fields of high homogeneity, arrangements of coils have been devised throughout which one and the same current of i amperes is sent. The magnetic field near the centre of the coils depends only on i and on the geometrical arrangements. The current i is measured by galvanometers or (more accurately) by null methods, in which a definite fraction of i is compensated by a Weston normal cell.

The Helmholtz-Gaugain system, mostly used, consists of two coaxial circular rings of radius a at the same distance a apart. Nagaoka [93]

has calculated the rectangular components F_x, F_y (x axial, y transverse) at a point whose coordinates are x, y relative to the centre, and gives the following expressions for F_x and the ratio F_y/F_x:

$$F_x = \frac{32\pi}{5a\sqrt{5}}\frac{i}{10}\left[1-\frac{18}{125a^4}(8x^4-24x^2y^2+3y^4)...\right] = \frac{0{\cdot}899716i}{a}[1-...],$$

(60)

$$F_y/F_x = 0{\cdot}576xy(4x^2-3y^2)/a^4+.... \tag{61}$$

With a precision of 1 in 10^5, the magnetic field in a region within the limits $x = \pm 0{\cdot}11a$, $y = \pm 0{\cdot}14a$ can be regarded as uniform. A current

$$i = 1{\cdot}1121a \text{ amperes} \tag{62}$$

would produce a field of 1 gauss at the centre.

In the original Schuster-Smith magnetometer, now used at Abinger, where H is approximately $0{\cdot}18$ gauss, Smith [24] made each of the two circular coils of twelve turns of wire, and took $a = 30$ cm.; hence a current of $0{\cdot}5$ ampere produces a central intensity of $0{\cdot}18$ gauss. He estimated that the magnetometer (including the measurement of the current) is accurate to 4 parts in 100,000.

Fanselau [95] has calculated symmetrical arrangements of pairs of coaxial circular currents which provide a homogeneous field over a wider range than the Helmholtz-Gaugain type. For instance, symmetrical arrangements of two pairs of coils may be used. The rings of the outer pair may have the radius a_1, and the distance between their planes may be $2d_1$ (in a Helmholtz coil $2d_1 = a_1$), and the corresponding values for the inner pair may be a_2 and $2d_2$. Fanselau [95] and Braunbeck [95 a] find that the best results are obtained with this arrangement, if the relative sizes are

$$d_1/a_1 = 1{\cdot}10723, \; a_2/a_1 = 1{\cdot}30915, \; d_2/a_1 = 0{\cdot}36401, \; d_2/a_2 = 0{\cdot}27805. \quad (63)$$

A current of $0{\cdot}80852a_1$ amperes yields 1 gauss at the centre. The formulae corresponding to (60) and (61) then begin with terms of the eighth power in x and y.

Another coil aggregate constructed by Fanselau has the outer coils of radius $a_1 = 38$ cm., distance $2d_1 = 91$ cm., and the inner coils of radius $a_2 = 50$ cm., distance $2d_2 = 30$ cm. With eighteen turns in each of the four coils, $i = 1{\cdot}79$ amperes gives a central intensity of 1 gauss. The inhomogeneity of the field in the space within the inner coils is then less than 20γ; a Helmholtz pair, in order to provide similar conditions, would need to have a diameter of about 250 cm. This aggregate has proved useful for testing magnetometers, because it can generate

magnetic fields of the same intensity as those for which the magneto-meter is to be used.

Three such coil aggregates, set up concentrically, with axes towards the north, east, and vertical, will provide a space in which the magnetic field is nearly uniform, and in which the magnetic intensity and its direction can be changed in a measurable way; for instance, this space can be made free from magnetic force by annulling the earth's magnetic field. (See also [95 b].)

2.35. The general theory of variometers [52]. Consider a magnetic 'needle', that is, a magnet of axis A and moment M, which is free to turn about a fixed axis B perpendicular to A; this is the general case of a magnetic variometer. Suppose that the needle takes up a position of stable equilibrium under the influence of mechanical couples, namely, (i) the couple MS produced by the component S of the earth's field perpendicular to the plane through the axes A and B, (ii) the couple MF, produced by the field of stationary auxiliary magnets (called *deflectors*), and (iii) the couple G produced by mechanical forces (gravity, or the torsion of a suspending thread). In equilibrium

$$S+F+G/M = 0. \tag{64}$$

Let the direction of the magnet be reckoned from a standard direction, corresponding to that which the magnet takes up when S is equal to the adopted *base-value* S_0 of the variometer. Let F_0, G_0 be the values of F and G in this position of the needle. Let ϕ denote the angular displacement of the needle from this standard direction, when S has any other value, and let $\Delta S = S - S_0$; ϕ is supposed to be small, and to be reckoned in circular measure, so that we may take $\cos\phi = 1$, $\sin\phi = \phi$. In the new position of the needle, F and G may have values $F_0 + \delta F$ and $G_0 + \delta G$; it is supposed that M remains constant, and that we may write $\delta F = -F'\phi$, $\delta G = -G'\phi$. Here F' is the component of the auxiliary magnetic field along the standard direction. The magnetic couple in the new position is $M(S_0 + \Delta S)$, produced by the component of magnetic force transverse to the standard direction, plus a couple $MT_0\phi$ produced by the component T of the earth's magnetic field along the standard direction; we may here neglect ΔT and put T equal to an average value T_0, considering $M\Delta T\phi$ a second-order term. Therefore, in the new position, (64) becomes

$$S_0+\Delta S+F_0+G_0/M+\phi(T_0-F'-G'/M) = 0,$$

or, subtracting the same equation for $\Delta S = 0$ and (consequently) $\phi = 0$,

$$\Delta S = -\phi(T_0-F'-G'/M) = \epsilon\phi, \qquad \epsilon = -T_0+F'+G'/M. \tag{65}$$

Consider, for instance, a variometer for the north component X. The needle is directed from west to east, and is free to turn around a vertical axis (so that G' is due entirely to torsion); in this case $T_0 = -Y_0$, and the angular scale-value becomes

$$\epsilon = Y_0 + F' + G'/M.$$

By suitable choice of F' and G'/M, any scale-value ϵ may be obtained.

The *linear* scale-value on the record is

$$e = \epsilon/a \qquad (66)$$

if a is the length of the optical path; if there is only a single reflection, a is equal to the double distance from the recorder to the mirror in the variometer.

If the optical path is long enough, and if the beam of light reflected from the mirror attached to the magnet ordinarily moves only through a small angle from its standard direction, then the ordinates on the record are nearly proportional to the angle ϕ. For example, if the width of the photographic sheet is 20 cm. and $a = 300$ cm., the range of ϕ on either side of the central direction is $\pm \mathrm{arc}\ 1/30 = 1{\cdot}9°$, and

$$\sin \alpha = \alpha(1 - \alpha^2/6) = \alpha(1 - 1/5400).$$

A second-order term may, however, become appreciable for H- or X-variometers, where the mechanical couple G is produced by the torsion of a quartz-thread. If the torsion coefficient of the thread is τ and G is produced by twisting the upper end of the thread through an angle θ, then $G = \tau\theta$. If $\theta = \theta_0 + \phi$ and (disregarding F for the present) the couple $(X_0 + \Delta X)\cos\phi$ or $(X_0 + \Delta X)(1 - \tfrac{1}{2}\phi^2)$ is balanced by G/M, equal to $\tau(\theta_0 + \phi)/M$, then (since $X_0 = \tau\theta_0/M$)

$$\Delta X = (\tau/M)\phi + (\tau\theta_0/M)\tfrac{1}{2}\phi^2 = (\tau/M)\phi(1 + \tfrac{1}{2}\theta_0\phi). \qquad (67)$$

If $h = \phi a$ is the ordinate measured on the record and e_0 is the linear scale-value (τ/Ma) for small values of ϕ or h, we may write

$$\Delta X = e_0 h(1 + h\theta_0/2a). \qquad (68)$$

The actual scale therefore varies across the sheet and increases with ΔX. Near an ordinate h_1 (for instance, near the base-line) small changes of the ordinate, of amount x, correspond to force-changes $e_1 x$; if we put $h = h_1 + x$ in (68), we obtain, for small values of x (so that $(h_1 + x)^2 = h_1^2 + 2h_1 x$), $\Delta X = e_0(h_1 + h_1^2\theta_0/2a) + e_1 x$, where

$$e_1 = e_0(1 + h_1\theta_0/a). \qquad (69)$$

For small values of x a constant linear scale-value e_1 can therefore be taken, and that is the usual observatory practice even for higher values of x. An 'auxiliary base-line' is taken near the ordinary position

of the spot; and for large disturbances the accuracy of the value of ΔX computed from (69) is quite sufficient because in such cases the absolute accuracy needed is not so high. But for observatories like Huancayo, where H varies daily across the whole sheet, the full formula (68) must be used; there, for instance, $e_0 = 2\cdot20\,\gamma/\text{mm}$. near the base-line, and $e_1 = e_0 + 0\cdot004h_1$ at h_1 mm. distance from the base-line, so that $e_1 = 3\cdot00$ for $h_1 = 200$ mm. Comparing (68) and (69), we find then, for the force variation ΔH corresponding to a deviation h_1 mm. from the base-line, $\Delta H = 2\cdot20h_1(1 + 0\cdot002h_1)$.

The scale-values are best determined by applying artificial fields produced by means of Helmholtz coils.

Various devices have been introduced to simplify the conversion of the readings h_1 into values of X, as, for instance, glass-scales suitably graded and marked so that mean hourly values of X can be read directly [69].

Great care must be taken that the orientation of the magnet is correctly perpendicular to the force component to be measured. At some stations the declination changes by as much as 2 degrees in ten years, and it is important to keep the variometers adjusted. The magnet of the Z-variometer must be horizontal, and its azimuth must be at right angles to the declination, and it must turn about a horizontal axis directed along the declination, because otherwise variations of the horizontal components affect the Z records.

Except in the case of D-variometers, which are pure direction-recorders, it is necessary to take account of any changes in the temperature of the instrument, owing to temperature-effects on M, F, and G. If t_0 is a standard temperature, the modified reduction formula is

$$\Delta S = \epsilon\phi + \alpha(t - t_0); \tag{70}$$

α is determined experimentally by heating and cooling the room, while a variometer in another room is kept at a constant temperature. In the expression (65) for ϵ, T_0 is independent of the temperature t, but as t increases, the field F' of the deflectors decreases, whereas G'/M increases. Hence, by suitable choice of F' and G'/M, $F' + G'/M$, and therefore ΔS, can be made independent of the temperature t. Such devices for *temperature compensation* are valuable, if care is taken that there are no appreciable temperature-differences between the different parts of the variometer.

La Cour [56] has introduced an ingenious optical temperature compensation: the recording beam of light passes through a prism p_1 attached

to a bimetallic support l, adjusted so that the motion of the prism, due to temperature-changes, corrects the deviation of the beam due to temperature-influences on the variometer itself (magnet C, with mirror B attached) (Fig. 20).
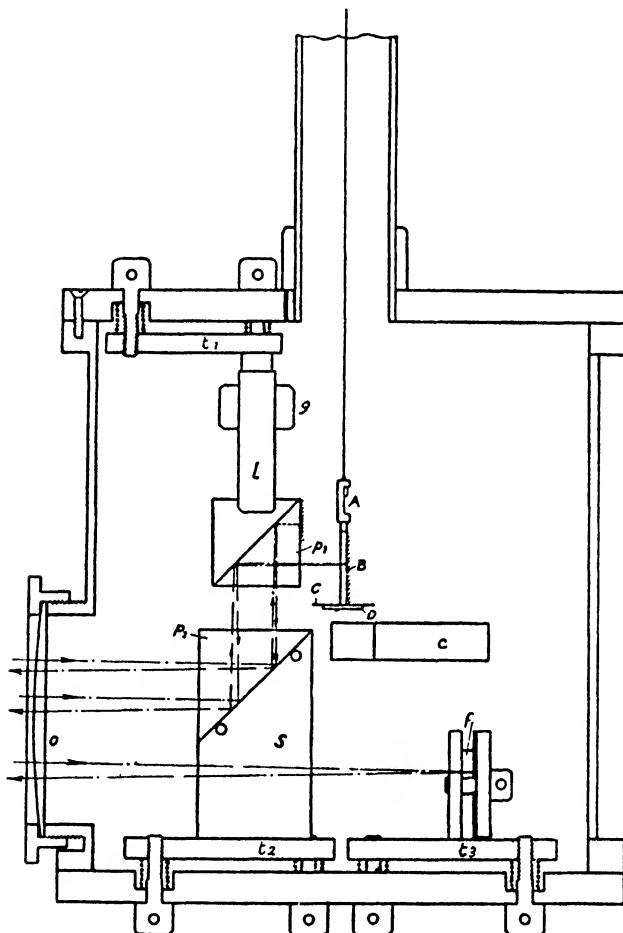
FIG. 20. Horizontal-intensity variometer with optical temperature compensation, by La Cour

Another possible method of temperature compensation is by the use of a 'quasi-astatic' magnet system for the needle of the H-variometer; i.e. two magnets of different moments and different temperature coefficients, in anti-parallel positions in a frame, adjusted so that a magnetic moment remains for the system, but no temperature-effect; mention may also be made of the temperature compensation in the Askania field variometers (Z-balances) by the use of different metals;

Heiland has discussed the theory of this type of compensating device (see 4.3).

The safest plan will always be to maintain a constant temperature in the room in which the variometers are housed; the variometers should certainly be preserved from daily variations of temperature; the slow annual variation is less dangerous, because most of its effect can be adjusted in the base-line values.

2.36. Types of variometers. The suspension of the magnet in a D-variometer is unifilar, with a thin thread of small torsion. If the scale-value of the D-variometer is to be adjusted to some fixed value, artificial fields can be introduced, or a strong fibre may be used and the magnet turned through 180°. In the H-, X-, and Y-variometers the suspensions may be bifilar or unifilar; in the latter case—now mostly used—a strong quartz thread is convenient, because such threads are free from elastic after-effects. No temperature compensation is needed in the simple type of D-variometer; in the other instruments it may be provided by deflector magnets or by La Cour's prism (§ 35); in these variometers the axis of the magnet is perpendicular to the component to be measured, and turns around an axis perpendicular to both directions.

La Cour's H-variometer has a very small magnet of cobalt steel (weight 0·025 gm., length 8 mm., moment 0·8 gauss cm.³); the (single) quartz thread is manufactured with conical and cylindrical ends of quartz for fastening. The temperature compensation is optical [55].

The Z-variometer is usually of the type known as Lloyd's balance, namely, a horizontal magnet balanced on a knife-edge, with the magnetic axis perpendicular to the direction of D. The couple ZM exerted by Z is compensated by a couple G exerted by gravity g. If W is the mass of the magnet, $m\ (= M/W)$ its specific magnetization, and s the horizontal projection of the distance between the mass-centre and the knife-edge, then

$$G = gWs = ZM = ZWm,$$

whence $$s = mZ/g. \qquad (71)$$

Hence s is independent of the mass (or size) of the magnet. For $m = 100$ gauss cm.³/gm., $Z = 0·4$ gauss, $g = 1,000$ dynes/gm., s is only 0·04 cm., and a shift of s by only 10^{-6} cm. would affect the readings as much as would a change of Z by 1γ. Every change in the relative position of the magnet and the knife-edge will result in slow shifts or sudden 'jumps' of the base-value, which must not be mistaken for real changes of Z.

In La Cour's Godhavn Z-balance the magnet, mirror, and knife-edge form a single piece, made of tungsten steel of 2·5 gm. weight, and length 5 cm.; this is hermetically sealed in rarefied air, and optically compensated for temperature.

Variometers to register the inclination I are also of the knife-edge type, with the magnet perpendicular to the direction of the earth's force and swinging in the plane of the magnetic meridian.

The troublesome knife-edge support has been replaced by thread suspensions in several balances (Watson [60], Angenheister [61], and others).

Another method of measuring Z, aiming at the elimination of the knife-edge, consists in converting the force-changes ΔZ into horizontal field-changes which can be measured by means of horizontal needles turning around a vertical suspension. For this purpose McNish [68], who mentions earlier attempts, uses vertical bars of perminvar (45 per cent. nickel, 25 per cent. cobalt, 30 per cent. iron) which has been subjected to a special heat treatment; for intensities of the order of the earth's field it has the permeability 300, and negligible remanent magnetization (for example, after a transient field-increase from 0·540 to 0·554 gauss, less than 1/10,000 of the additional induced magnetization remains). Four bars (each $5 \times 5 \times 52$ mm.) of this material (F in Fig. 21) are mounted vertically in pairs on pyrex glass rods, one above the other, leaving an air-gap of about 12 mm. Two smaller strips A ($2 \times 2 \times 7$ mm.) are mounted vertically on the sides of a mirror M serving as armature, which is suspended by a quartz fibre so as to hang in the centre of the air-gap. The torque exerted by the torsion of the quartz fibre is counteracted by the action of the induced magnetic moments of the rods, which produce an inhomogeneous field and draw the armature into the position of maximum magnetic flux (see plan). The performance of this instrument (Fig. 21 and Plate 12 a) has been satisfactory; it reacts only to changes in Z if suitably adjusted.

2.37. The quartz horizontal-force magnetometer (QHM). In Lamont's theodolite (§ 4) the position of the needle for a reading is always such that the optical axis of the theodolite is perpendicular to the mirror attached to the needle. Suppose that the needle, of moment M, is suspended by a strong quartz thread of torsion coefficient τ. Under the couple $MH \sin \alpha$ due to H, and the torsional couple $\tau\delta$ (where δ denotes the angle of twist through which the upper end of the thread is turned relative to the lower end), the needle will take up a position of equilibrium (P). The azimuth α of the normal to the mirror in this

PLATE 12
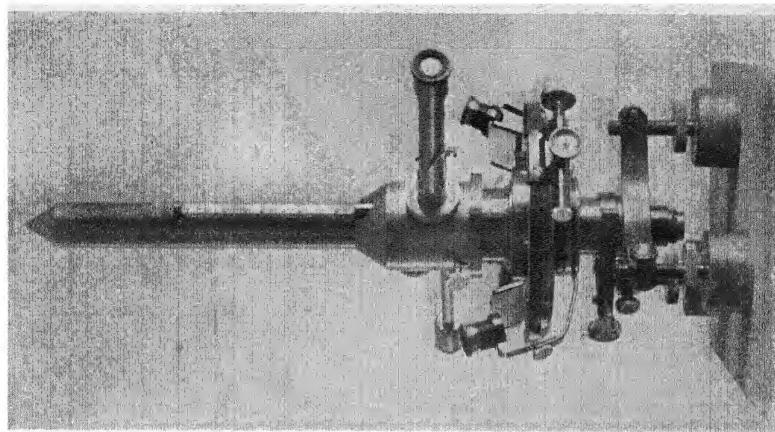
b. The **QHM** (quartz horizontal magneto-meter) mounted on a theodolite-stand (La Cour)
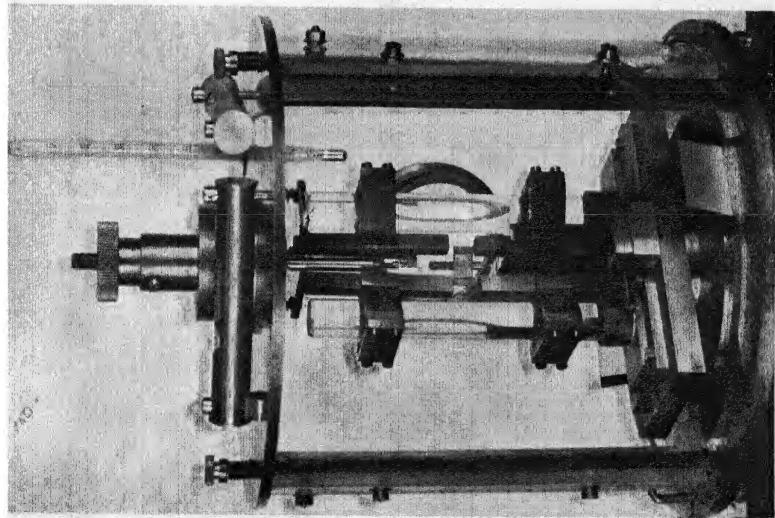
a. Vertical-force induction variometer (McNish)

*
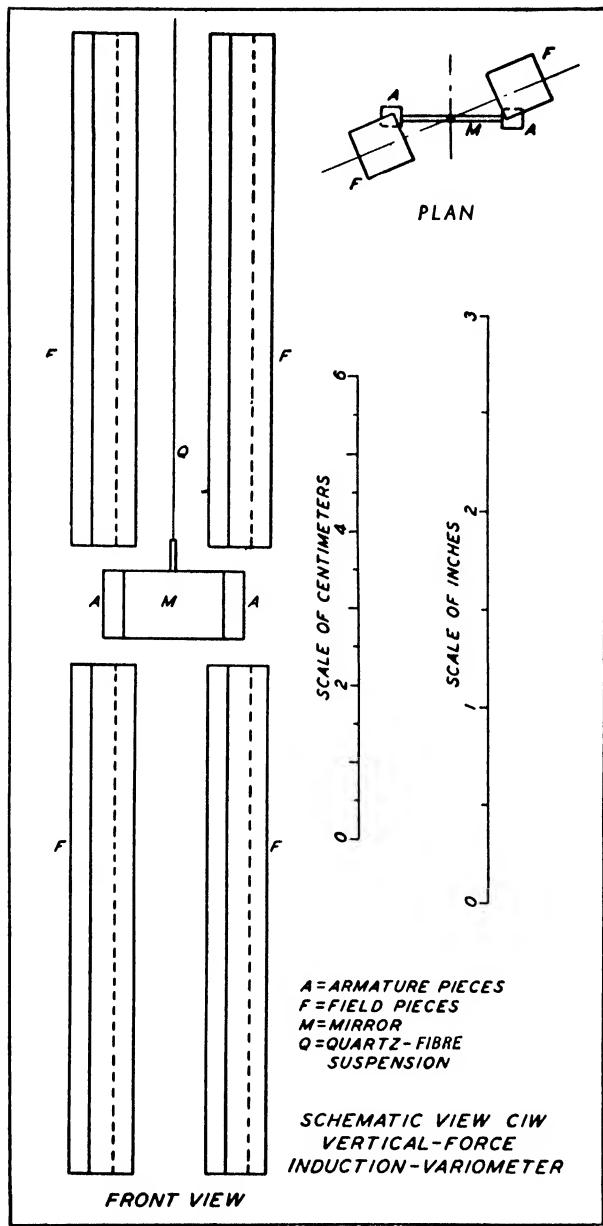
FIG. 21.   Schematic elevation and plan of the McNish
vertical-force induction variometer

position P, reckoned from the magnetic meridian, is read from the
divided circle B of the theodolite. If the theodolite is given a full turn
(2π) towards the right, the torsion of the quartz fibre is increased in

the same direction, so that the magnet will also be turned, by a small angle, to the right; hence, to make the optical axis of the telescope perpendicular to the mirror, the theodolite must be turned through an additional angle a_1 corresponding to a position P_1 of the needle; a_1 is obtained as the difference between the azimuth readings at P and P_1. In the position P_1 the twist of the thread is exactly $\delta+2\pi$. The theodolite is next turned through two full turns towards the left, and slightly farther into another position of equilibrium P_2 of the needle; the azimuth is again read and differs by an angle a_2 (say) from that of P; the twist of the thread in the position P_2 is then exactly $\delta-2\pi$. The three equations of equilibrium are

$$MH\sin\alpha = \tau\delta, \qquad MH\sin(\alpha+a_1) = \tau(\delta+2\pi),$$
$$MH\sin(\alpha-a_2) = \tau(\delta-2\pi). \tag{72}$$

Adding the last two equations and eliminating δ by means of the first, we obtain

$$\tan\alpha = (\sin a_1-\sin a_2)/(2-\cos a_1-\cos a_2); \tag{73}$$

also, by subtraction of the last two equations of (72), we find that

$$H = (4\pi\tau/M)/[\sin(\alpha+a_1)-\sin(\alpha-a_2)]. \tag{74}$$

If the instrument is well adjusted, so that δ and α are small, then a_1 and a_2 (corrected for time-changes of declination) differ little from their mean value $\phi = \frac{1}{2}(a_1+a_2)$, and

$$H = (2\pi\tau/M)/\sin\phi. \tag{75}$$

If M and τ are assumed constant—apart from linear changes due to temperature, and the additional moment induced in the needle by H—then H is directly measured by ϕ; the excellent elastic properties of quartz fibres (tested by Venske [98]), and the slow change of the moment of suitably treated cobalt magnets, warrant this assumption.

These considerations guided La Cour [96, 97] in the construction of his quartz horizontal magnetometer (QHM). This is nothing but a suspension tube enclosing a short magnet (of 1·5 cm. length and 2 gauss cm.^{3} moment) suspended by a quartz fibre. A small telescope is part of the instrument, which can be mounted on the base of any theodolite so that readings can be taken on the divided circle.

The actual working formula used is

$$\log H = C-\log\sin\phi+c_1 t-c_2 H\cos\phi, \tag{76}$$

where t denotes the temperature, c_1 the temperature coefficient, and c_2 the induction coefficient (§§ 4, 31). For the QHM instrument no. 11, for instance, the numerical values are as follows:

$$\log H = 0·18181-1-\log\sin\phi+0·000150\,t-0·0022\,H\cos\phi. \tag{77}$$

The instrument is shown in Plate 12 b, mounted on a theodolite base ready for observation. The satisfactory working of the instrument is mainly due to the new mounting device for quartz fibres introduced by La Cour (Fig. 22). The usual clamping of the fibres did not prevent slipping, and injured their upper ends. La Cour leaves a drop of quartz at the end and places this in the holder shown in Fig. 22; subsequent light heating fuses the quartz and provides a safe connexion with the holder.

The QHM instrument is used for comparisons of magnetic standards and also for other purposes such as local surveys, tests for the homogeneity of the field in variometer rooms, etc. It is so robust and light (weight 500 gm.) that it can be sent by mail—preferably three instruments in the same box, to ensure against changes in the magnetic moment M of one instrument. The measurements take only a short time, and the standard of accuracy is very high, well within 1γ. The accuracy is increased by working with large deflexions ϕ of the order of $60°$.

FIG. 22. Three views of the suspension device for the upper end of the quartz fibre in La Cour's quartz horizontal magnetometer (QHM). The suspension used at the lower end is seen at A in Fig. 20

2.38. The magnetron. The magnetron, devised by Greinacher and Hull, is a vacuum tube with a cathode in the shape of a straight filament, heated by a current, situated at the axis of a cylindrical anode; the filament serves as a source of electrons. The current i between the filament and the plate can be practically stopped by applying a magnetic field R parallel to the filament; the paths of the electrons leaving the filament are then bent in circles by the magnetic field, and so the electrons do not reach the anode (see 15.10). Below a certain critical value of R, i is practically independent of R, but near the critical value i decreases rapidly to zero; a change of 20γ corresponds, for instance, to a current change of 2×10^{-7} ampere.

M. Rössiger [100] places the magnetron inside a coil which produces a constant auxiliary field R_0 near the critical value. The magnetron with this coil forms the indicator. The indicator is mounted rigidly inside a pair of Helmholtz coils, and the whole system may be turned about a horizontal and vertical axis. The anode current is regulated

by the component of the earth's field parallel to the filament, by the constant auxiliary field R_0, and by the field of the Helmholtz coil.

In order to measure the horizontal intensity H, the filament is brought into the horizontal position. The apparatus is rotated about a vertical axis, and the anode current is read from an ammeter; its maximum and minimum readings indicate when the filament is in the magnetic meridian. Having fixed the filament in the magnetic meridian, the anode current is read accurately; then the instrument is turned through 180° around a horizontal axis, and the anode current is brought back to its initial value by sending a current through the Helmholtz coils. H is then twice the compensating field of the Helmholtz coil. Similarly Z or any other component may be measured.

The chief sources of inaccuracy with this instrument are a bad vacuum (which can be remedied by thorough baking) and fluctuations in the filament current, which can be eliminated by a compensation device. The accuracy of the method is at present 40γ. Its chief advantage for measurements on shipboard is the absence of moving parts and therefore of dynamic deviations (p. 46).

2.39. Magnetic gradiometers. It has sometimes been suggested that in geophysical prospecting (Ch. IV) it will be useful to measure the space gradients of the magnetic field, that is, the nine quantities

$$\partial X/\partial x, \qquad \partial X/\partial y, \qquad \partial X/\partial z,$$
$$\partial Y/\partial x, \qquad \partial Y/\partial y, \qquad \partial Y/\partial z,$$
$$\partial Z/\partial x, \qquad \partial Z/\partial y, \qquad \partial Z/\partial z.$$

Only five of these are independent, since they include three equal pairs ($\partial X/\partial y = \partial Y/\partial x$, $\partial X/\partial z = \partial Z/\partial x$, $\partial Y/\partial z = \partial Z/\partial y$, because X, Y, and Z are the derivatives of a potential, $X = -\partial V/\partial x$, etc.), and also because of Laplace's equation (1 (14)) $\partial^2 V/\partial x^2 + \partial^2 V/\partial y^2 + \partial^2 V/\partial z^2 = 0$, or

$$\partial X/\partial x + \partial Y/\partial y + \partial Z/\partial z = 0. \tag{78}$$

The order of magnitude of each of these three terms is $10^{-4}\gamma/$cm. in the undisturbed field of the earth, and may reach values 10^5 times higher in local anomalies; but the sum of the three terms must reduce to a value of $10^{-7}\gamma/$cm., which is the limit if the change in the magnetic susceptibility of air with height is taken into account.

A magnetic gradiometer, constructed by Roman and Sermon for prospecting purposes [102], consists of two similar induction coils rotating about parallel shafts in such a manner that the planes of the coils always remain parallel. For instance, if the centres of the coils are at the ends of a line orientated along the magnetic meridian, and if the

coils are rotated about parallel horizontal axes, the difference between the currents induced is proportional to the gradient of Z along the meridian. The instrument mentioned measures the ratio of the two electromotive forces by a specially devised Kirchhoff net. Chapman [103] has discussed the theory of such instruments, also for measuring the time rate of change of the space gradients in polar regions.

Berroth and Schleusener [103a] have discussed the possibility of measuring $\partial Z/\partial x$ and $\partial Z/\partial y$ by means of a torsion balance, in which the lower mass is replaced by a vertical iron bar.

2.40. Non-magnetic material for magnetic instruments. Even the slightest trace of magnetic material present in magnetic instruments is liable to distort the magnetic field lines and give wrong results. This has not been realized sufficiently by instrument-makers, so that many instruments are affected by serious errors. For the Department of Terrestrial Magnetism, Carnegie Institution of Washington, Huff and Steiner have developed the following method of producing non-magnetic copper and brass castings [103c].

Clean copper is placed in a crucible with a quantity of good non-ferrous foundry flux, the initial charge of flux being one-third of the total quantity. A layer of powdered charcoal is added, and when the first charge has melted, more copper (preferably pre-heated) is added with another one-third of the flux. After all the copper has melted, zinc is added as a deoxidizer (in the proportion 1·5 pounds of zinc to 100 pounds of copper) and is well stirred in. At this point the metal should be thoroughly skimmed. The last portion of flux is then added and well stirred in the melt. Throughout the melting it is important that the copper be kept at a low temperature near the melting-point and at no time be allowed to 'boil'. Just before the crucible is taken from the fire a thin blanket of pure silica sand is spread over the molten copper and allowed to melt, thus forming a glassy blanket which retains the impurities remaining on the surface of the metal. When pouring (at the lowest temperature at which the metal will run) this glassy blanket can be easily held back with a skimmer.

In order to be able to keep the metal at the lowest pouring temperature it is necessary to design the patterns so that the mould can be filled rapidly. This is done by eliminating all very thin walls, fins, and ribs.

Stock brass screws and fittings are usually magnetically impure. Each part of a magnetic measuring instrument, and the whole instrument, when assembled, should be tested thoroughly, in order to detect, for instance, steel splinters from tools. Johnson and Steiner [103b] have developed an adequately sensitive astatic magnetometer for this purpose (see pp. 142, 3).

THE EARTH'S MAIN FIELD AND ITS SECULAR VARIATION

3.1. Isomagnetic charts. The results of magnetic surveys reduced to a common epoch (2.8) can be conveniently represented graphically by charts of various kinds. One method is to draw lines having the property that a given magnetic element has a constant value along each such line. The lines are called *isomagnetic lines*, and a chart in which the distribution of a magnetic element is thus indicated, whether for the whole or a part of the earth, is called an *isomagnetic chart*. Special names are assigned to the lines and charts for certain of the elements: those referring to the magnetic elements D, I, and X, Y, Z, H or F respectively are called *isogonic* (D; *agonic* for $D = 0°$), *isoclinic* (I), and *isodynamic*. The charts may be drawn on a globe or on a plane projection of any kind.

For nautical and aeronautical purposes, and for use in land and mine surveying, the isogonic chart is the most important, and several national hydrographic or other offices publish such charts from time to time [G 21–8]; less frequently other isomagnetic charts are produced. Figs. 1–4 give reproductions, on a greatly reduced scale, of world isomagnetic charts for various elements and epochs, namely for D (for 1935 and 1922), and for H, I, and F for 1922. They are largely based on the survey work of the Carnegie Institution of Washington by land and sea.

The usual Mercator projection (Fig. 1) is commonly chosen as the base-map for world charts, because it does not distort the azimuths of horizontal directions and is therefore suitable for navigation. However, it over-emphasizes the polar regions; e.g. an area near 60° latitude appears four times as large on the chart as an area of equal size near the equator. Projections giving equal areas (as used in Figs. 13–17) necessarily have curved meridians, which might give a wrong impression in vectorial charts (Fig. 8). Therefore a simple rectangular net has been chosen for a number of maps (Figs. 1 a–4), with a ratio of one degree of longitude to one of latitude $= 4:5$, so that there is no distortion near latitude 37°, while the equatorial belt appears slightly compressed and the polar zones appear stretched in the direction east–west.

Isomagnetic charts enable the value of the magnetic force at the epoch of the charts to be inferred, by interpolation, for any point of
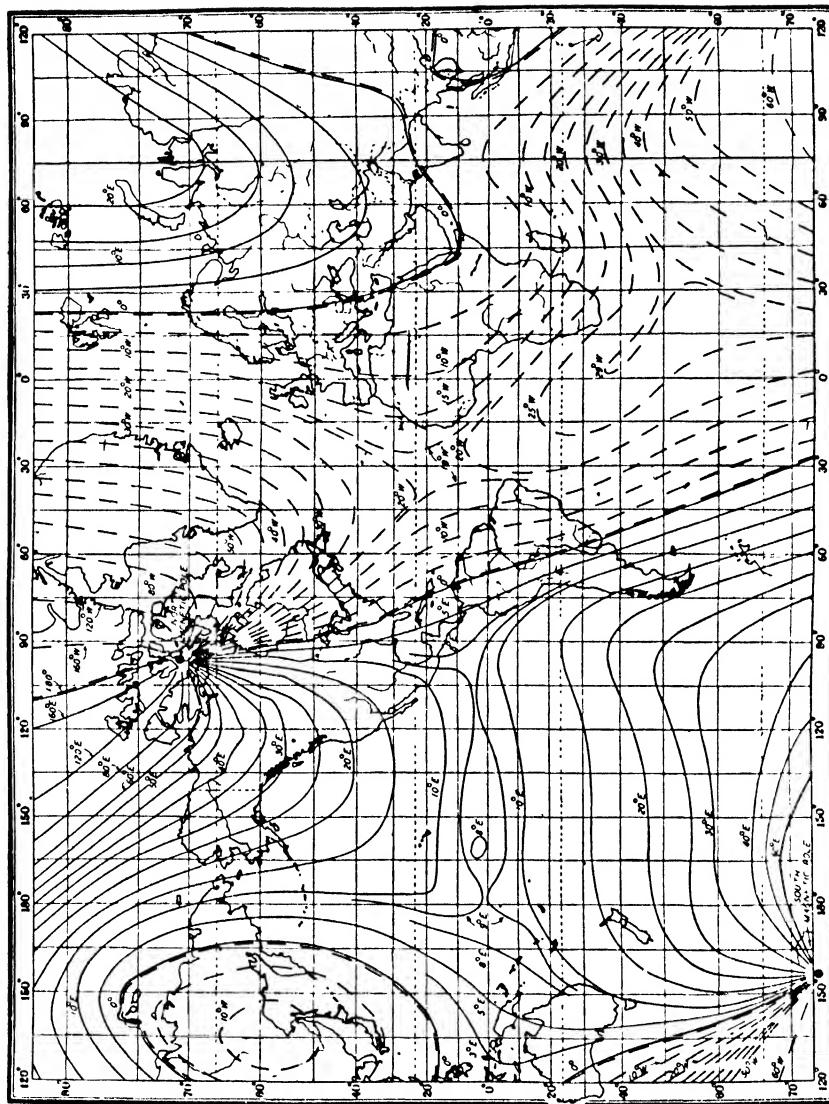
FIG. 1. World isogonic chart showing lines of equal magnetic declination, epoch 1935. (U.S. Hydrographic Office)
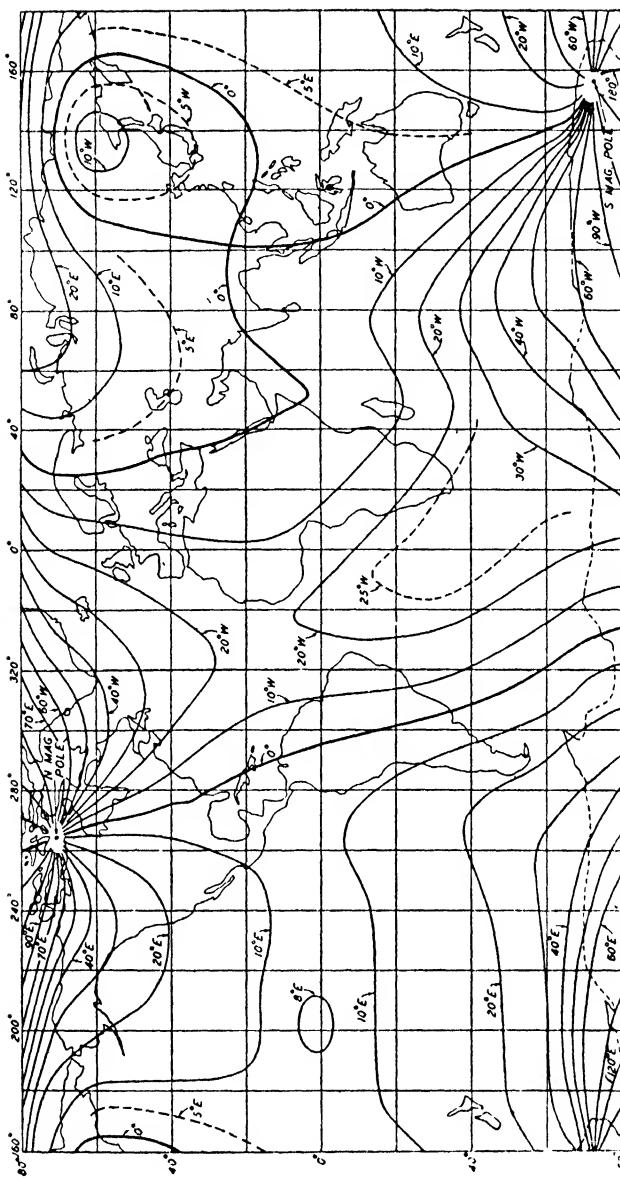
Fig. 1 a. Isogonic lines for 1922
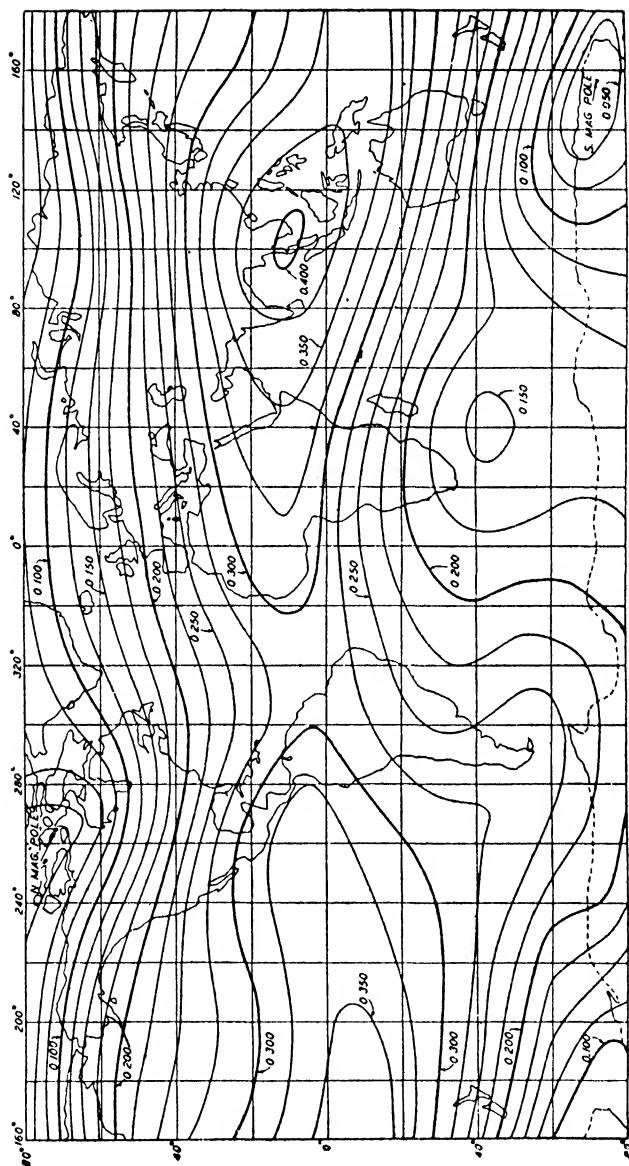
Fig. 2. Lines of equal horizontal intensity, 1922
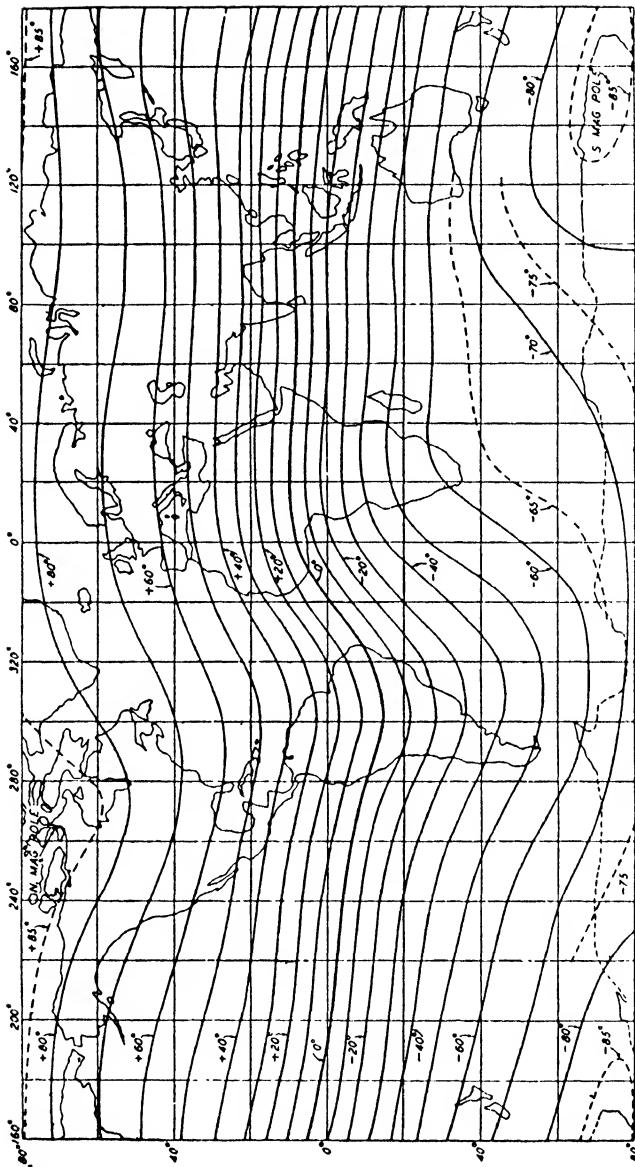
Fig. 3. Lines of equal inclination (*I*) for 1922
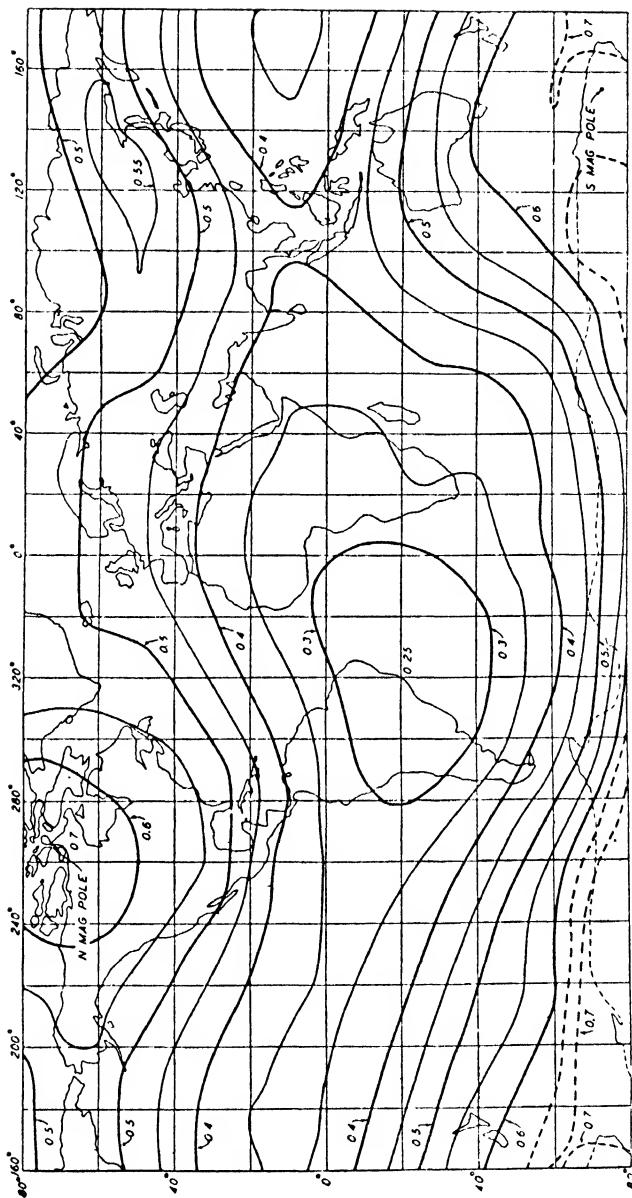
FIG. 4. Lines of equal total intensity (*F*) for 1922. (Drawn by Ennis from data scaled by Dyson and Furner from British Admiralty charts)

the earth's surface. The value so obtained may, however, be in error to some extent owing to local irregularities of the field on too small a scale to be represented on the chart: the more detailed a chart (and the survey on which it is based), the more numerous are the detailed irregularities (§ 5, and Ch. IV) brought to light. A world chart on any practicable scale cannot be expected to exhibit these small inequalities: its function is to give a 'smoothed' or general indication of the distribution of the earth's field. Even in this respect it is liable to be faulty if it is 'brought forward' from surveys made at an epoch many years before, by application of corrections for the secular variation, also determined some years previously. The Carnegie ocean surveys (2.7) revealed errors in the isogonic charts prior to 1922 which amounted to 2° or 3° in some regions, though the charts were in the main accurate in the more frequented regions of the earth.

A list of maps and surveys for individual areas is given in the bibliography [33–70].

3.2. Lines of horizontal force. The direction of the horizontal component of F can be indicated more readily to the eye and the mind by a chart showing the 'lines of horizontal magnetic force' than by the isogonic chart, though the latter is the more useful for the purpose of reading off the declination at any point. With a little trouble, however, either chart can be constructed from the other. The lines of horizontal magnetic force, or magnetic meridians, are defined by the property that the tangent at any point is along the direction of H. Fig. 5 is a reduced copy of such a chart for one hemisphere, prepared for the epoch 1830. It well shows the direction of H by the lines of force converging towards the magnetic north pole marked A.

The relative nature (1.2) of the declination D is manifested by the convergence of the isogonic lines to foci at the geographic as well as at the magnetic poles, while the 'intrinsic' magnetic meridians (p. 9) converge only towards the magnetic poles and show no particular features near the geographic poles (Fig. 6). The magnetic meridians are *not* lines of force in the ordinary sense (1.8), but projections of such lines; the interspacing of the meridians has therefore no immediate physical meaning.

It is unfortunately difficult to represent graphically, by a single chart, stereogram (2.19), or model, the distribution of F as a whole, in direction and magnitude. The mind must always be called on to combine data from more than one diagram in order to form a complete picture.

3.3. The surface field of a uniformly magnetized sphere. The earth's field F, as measured near the surface, can be described roughly as resembling that of a dipole, or that of a uniformly magnetized sphere (1.15). In such a sphere let AB be the diameter along the direction of magnetization. This is called the magnetic axis, A and B being the magnetic poles. The lines of horizontal (i.e. tangential) magnetic force
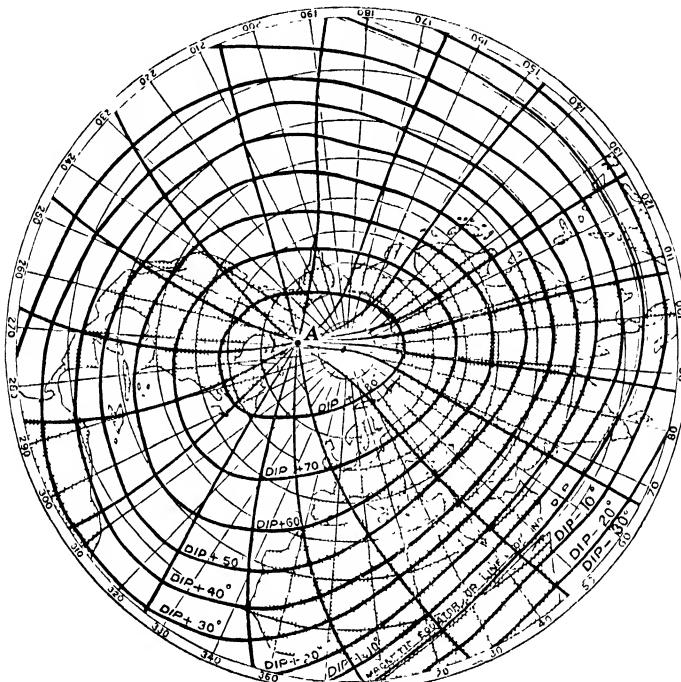
FIG. 5. Lines of horizontal magnetic force, and lines of equal inclination, for the northern hemisphere, epoch 1830. (After Airy [G 9])

H are meridian circles through AB. At any point P on the surface, at the angular distance θ from B,

$$H = H_0 \sin\theta, \qquad Z = 2H_0 \cos\theta, \qquad \tan I = Z/H = 2\cot\theta, \quad (1)$$

where Z is the downward vertical (or inward radial) component of force, and H_0 is a constant. Thus H is zero at the poles, and elsewhere is everywhere of the same sign. It attains its maximum value H_0 at the *magnetic equator* $\theta = \frac{1}{2}\pi$, where $Z = 0$; Z is reversed on crossing the equator, and attains its maximum value $2H_0$ at the poles, where $H = 0$. The magnetic equator $\theta = \frac{1}{2}\pi$ is the line of no dip ($I = 0$ or $Z = 0$). The magnetic moment of the sphere is $H_0 a^3$, where a denotes the radius.

3.4. The general character of the earth's field. The earth's field does not exhibit these simple properties exactly. The lines of horizontal magnetic force (Figs. 5, 6) differ somewhat from great circles: indeed their focal points, where $H = 0$ and $I = \frac{1}{2}\pi$ (the magnetic poles
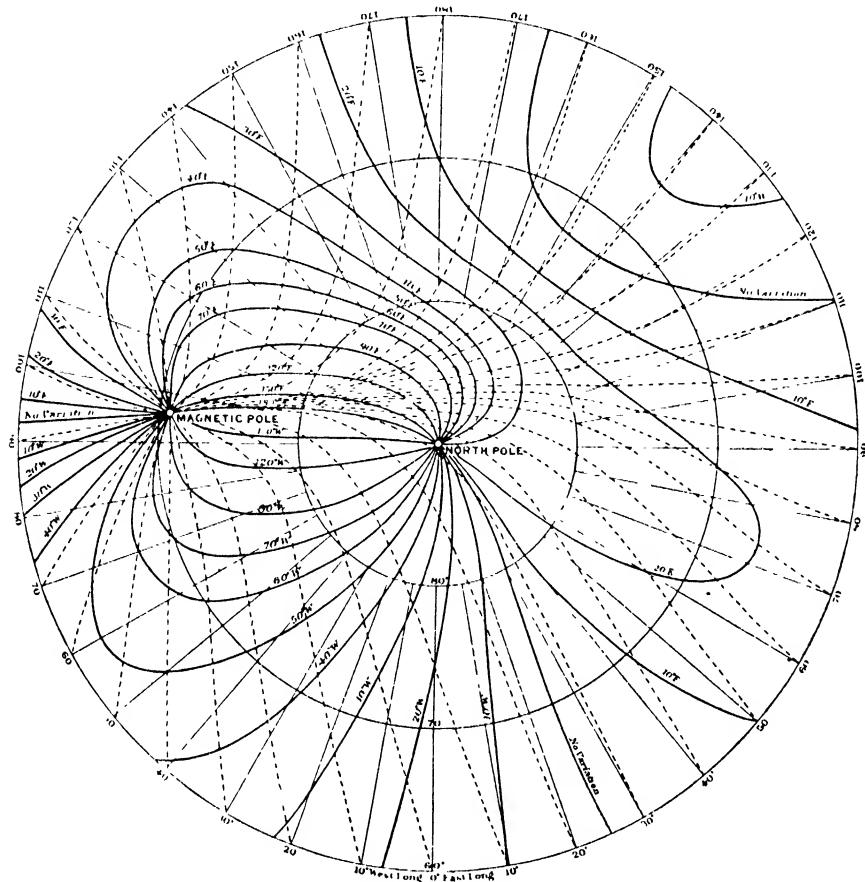
FIG. 6. Lines of equal declination (full lines) and lines of horizontal force or magnetic meridians (broken lines), epoch 1922, for the arctic region north of 60° latitude. (After H. Spencer Jones [3].) The agonic lines ($D = 0$) are marked 'No variation'

or poles of magnetic dip), are not quite antipodal. According to the 1922 Admiralty charts, the northern one (N') was at 71°N., 264°E. (96° 0' W.), and the southern one (S') at about 73°S. 156°E., each therefore being about 1,500 miles from the point antipodal to the other. Again, the line of no dip, or magnetic equator, is not truly circular (Fig. 3), nor is H constant along it. The average value of H along it in 1922 was 0·34 Γ, nearly half the polar values ($\pm 0·7$ Γ) of Z.

According to Bauer (18.3) the field of a uniformly magnetized sphere—equal in size to the earth—which would agree most closely with the earth's surface-field F in 1922 would correspond to the value

$$H_0 = 0 \cdot 311 \ \Gamma,$$

the pole B of the magnetic axis being at 78·5° N., 290·9° E. (or 69·1° W.). This *dipole field* (1.5) will be denoted by F_1^i or, usually, by F_1. The axis

SCALE IN C G S UNIT
0   0·2   0·4   0·6
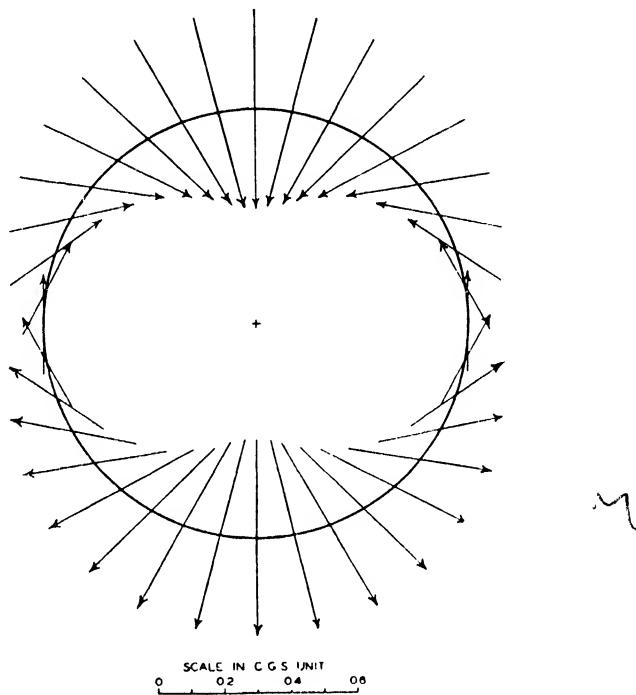
FIG. 7. Vectors showing the magnitude and direction of the magnetic force at the surface of a uniformly magnetized sphere, at different points (where the vectors cross the circle) along a meridian

AB and the poles A, B of this field F_1^i, and the diametral plane perpendicular to AB, may be called the *dipole axis*, the *axis poles*, and the *dipole (magnetic) equator*. The meridians through the dipole axis may be called the *dipole meridians*. Fig. 8 indicates the distribution of the horizontal force, and Fig. 9 the *dipole isogones*, for the dipole field F_1^i. The notation A (austral) and B (boreal) for the axis poles is due to Schmidt [G 10]; the pole B in the northern hemisphere has the same polarity as the south-seeking end of a magnetic needle (1.15).
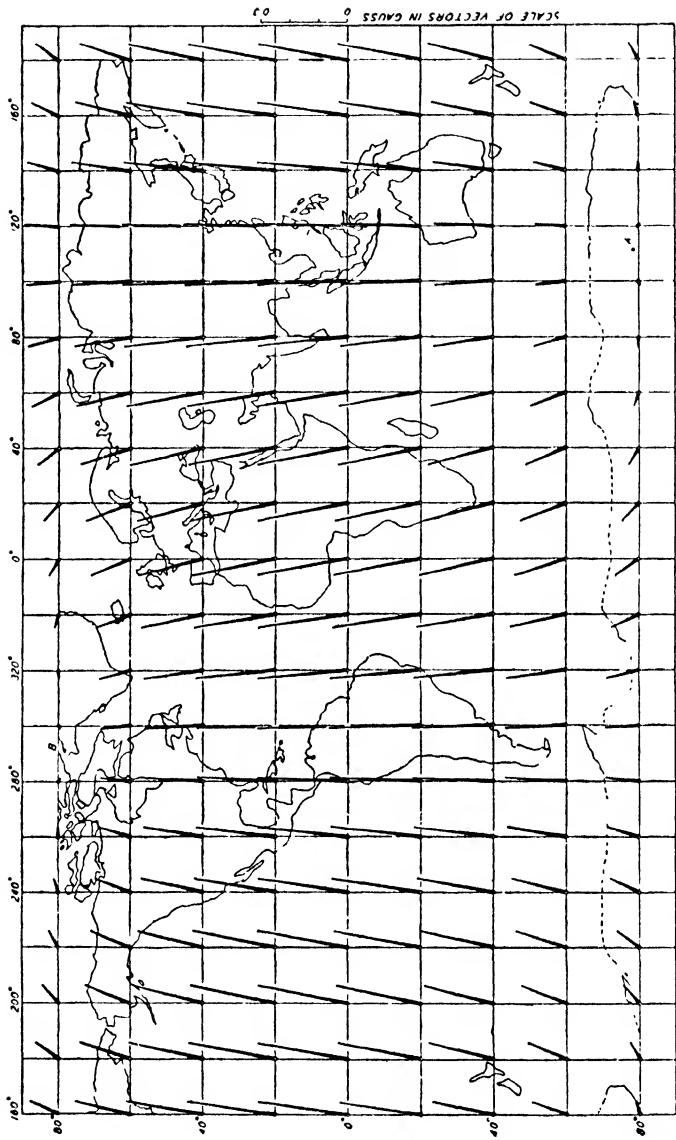
FIG. 8. Vectors showing at different points on the earth the magnitude and direction of the horizontal magnetic force corresponding to the field F_1 of the central dipole, or the field of uniform magnetization, that most nearly approximates to the earth's field F
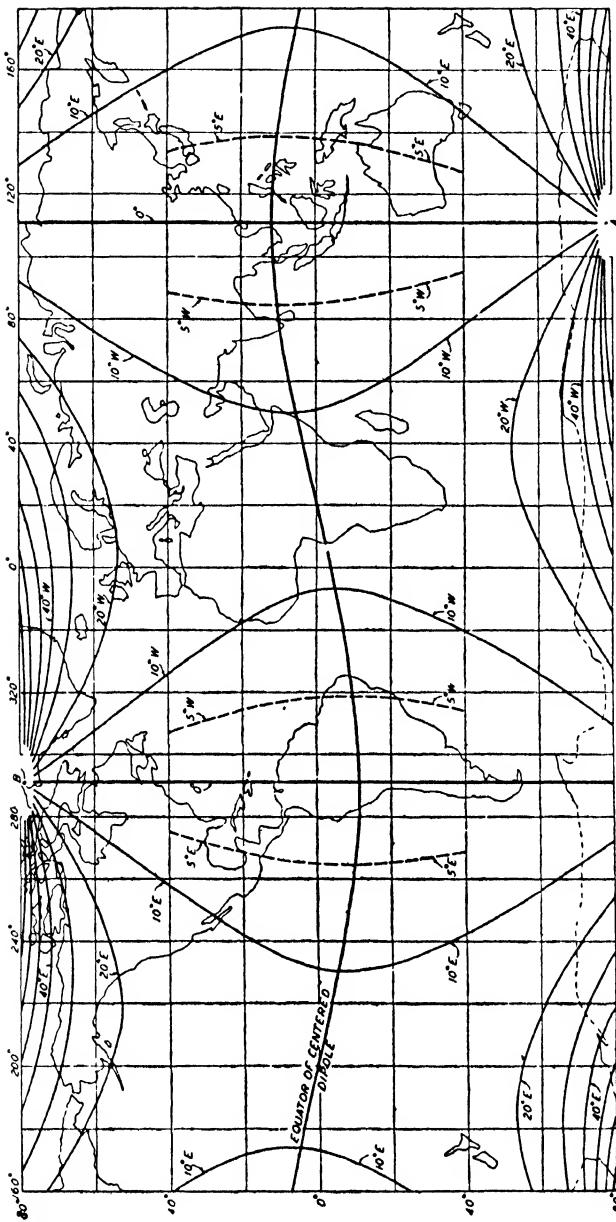
FIG. 9. Lines of equal declination corresponding to the field F_1 of the central dipole, or the field of uniform magnetization, that most nearly approximates to the earth's field F

3.5. Regional and local anomalies. The difference between the observed field F and the field F_1 is partly represented by the horizontal vector components of $F-F_1$ in Fig. 10, drawn on the same scale as those in Fig. 8; hence the vectors in Figs. 8 and 10, added, give F. The differences $F-F_1$ are clearly by no means small, and over great regions they are systematically in one direction, as, for instance, in middle Asia, in South Africa and the adjacent oceans, and in northern Canada. Between 70° N. and 60° S., where the field is known with some certainty, the average numerical values of the north, east, and vertical components of the differences are $3,200\gamma$, $2,800\gamma$, and $5,200\gamma$ respectively; the horizontal vector of the difference reaches the high value $11,000\gamma$ near South Africa. These *regional anomalies* account for the larger irregularities of the isomagnetic lines (Figs. 1–4).

Superposed on the regional anomalies there are *local anomalies* of much smaller extent; some of the most intense are limited to areas of a few hundred square miles, while others, due to a ferriferous rock magnetized by a lightning stroke, may cover only a few square feet. The greatest of these anomalies is that at Kursk, south of Moscow, where Z is raised from its normal local value $0.4\,\Gamma$ to $1.9\,\Gamma$ (cf. 4.11): the anomaly is confined mainly to two parallel strips running from NW. to SE., 40 miles apart; the principal strip is 160 miles long, and less than 2 miles broad in its most intense part. A similar long strip of high disturbance, where Z rises to $3.6\,\Gamma$, occurs in Lapland, and there are others of less note in Finland and elsewhere. In East Prussia the isogonic lines are found to be very complicated; instead of being nearly parallel, as in most regions, they divide up the district into a number of distinct areas separated by closed isogonic curves, 10 to 20 km. across, in which the anomalies of D may rise to 7°, corresponding to an anomaly of $2,300\gamma$ in the east component Y (Fig. 11).

Obviously a relatively small disturbing field superposed on F_1^i can considerably alter the position of the dip-poles N' and S', where H vanishes; strong local anomalies may even produce additional dip-poles. Hence these points are of less significance than the axis poles A and B determined by the main component, F_1^i, of F.

Local anomalies and their relations to geology will be discussed in Chapter IV.

Various ways of representing local anomalies on maps of different scales have been used, and the problem arises whenever smoothed isomagnetic lines are to be drawn from a set of original observations. It has been stated [1] that the density of magnetic stations is insufficient

SCALE OF VECTORS IN GAUSS

DIFFERENCES, OBSERVED MINUS CENTERED-DIPOLE FIELD, FOR HORIZONTAL VECTORS
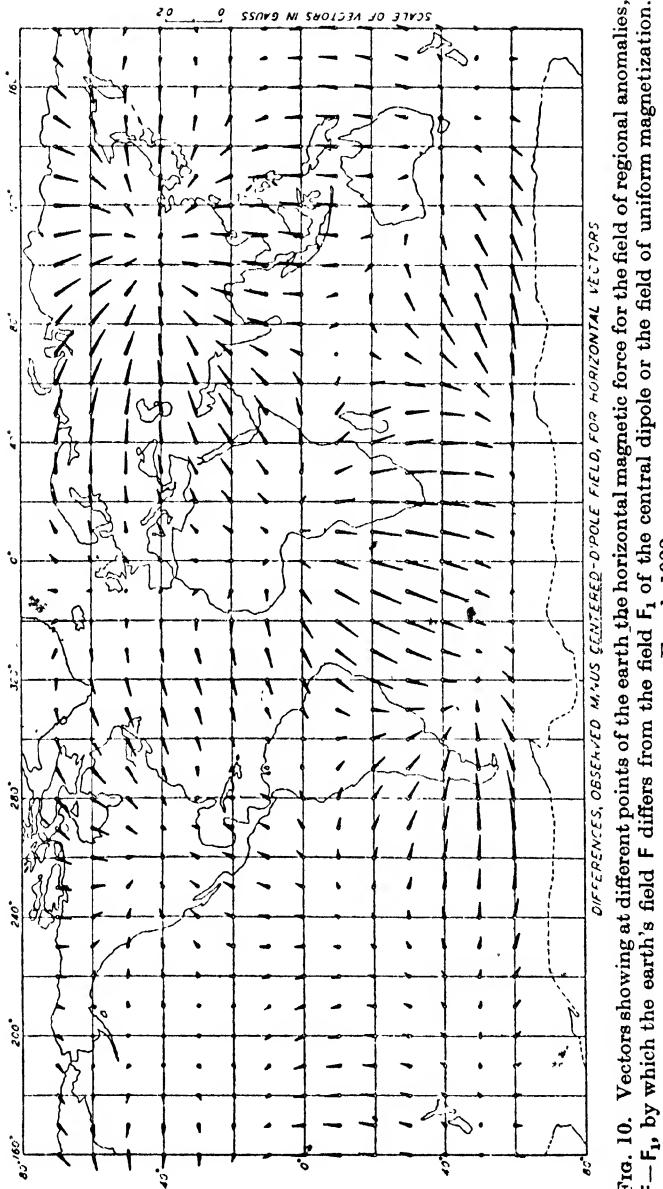
Fig. 10. Vectors showing at different points of the earth the horizontal magnetic force for the field of regional anomalies, $F - F_1$, by which the earth's field F differs from the field F_1 of the central dipole or the field of uniform magnetization. Epoch 1922

to represent the *true* distribution of magnetic lines in most regional magnetic surveys, where the observing stations are 10 miles or more apart. But even if the network of stations were much more dense, the true isomagnetic lines might not be easy to draw, or might be too complicated for convenient use. Smoothed isomagnetic lines are therefore recommended. The departures of the observed values from the
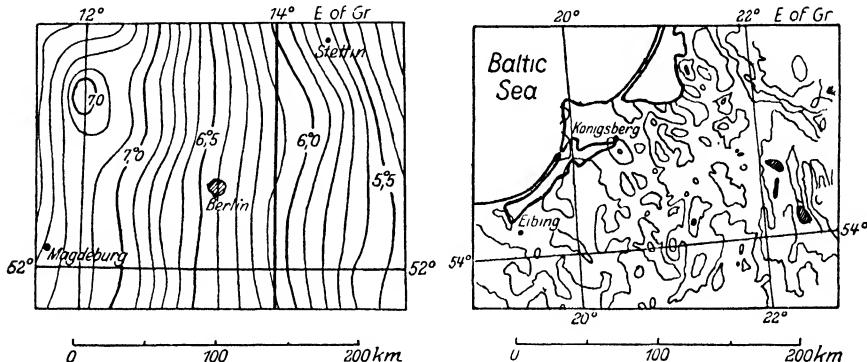
FIG. 11. Lines of equal declination for two regions in Germany (from Haussmann's chart [51] for 1925). Left: a region (near Berlin) which magnetically is relatively undisturbed; the isogones are drawn at 0·1° intervals, and the range of D is only 2·4° (from 5·1° to 7·5° W.). Right: a magnetically disturbed region in East Prussia; the isogones are drawn at 1·0° intervals, and the range of D is 7°; some small areas with eastern declination are hatched

smoothed field may be indicated by conventional symbols. An example of such a chart [2] is given in Fig. 12.

3.6. Line integrals of horizontal magnetic force. In so far as the earth's surface magnetic field F is due to magnetic matter within the earth, and/or to electric currents above or below the earth, but not situated at or traversing the earth's surface, the magnetic force F can be derived from a scalar potential V (satisfying Laplace's equation 1 (14)), of which it is the negative gradient (1 (30)). The representation of F in this way is of great importance, as Gauss showed, because by expressing V as a series of spherical harmonic terms it is possible to distinguish between the parts of F which arise respectively above and below the earth's surface (Ch. XVIII). If part of F is due to electric currents flowing near the earth's surface, their existence can be inferred by calculating the line integral (1 (24)) of F round any closed curve, which will give 4π times the current threading the curve, i.e.

$$\oint F \cdot ds = 4\pi \int i \cdot dS, \tag{2}$$

where ds, dS denote elements of the curve, and of any surface of which
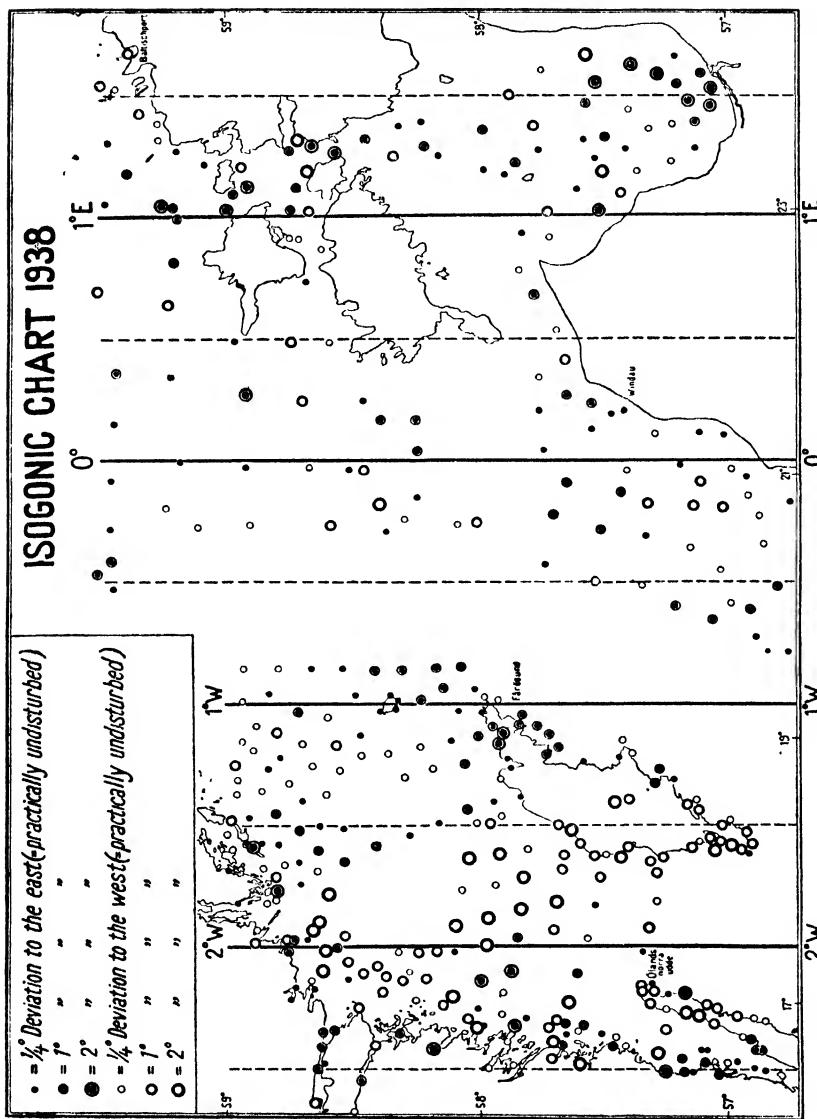
Fɪɢ. 12. Magnetic declination over the Baltic Sea, 1938. The 'normal' declination is indicated by smoothed isogones, and the departures from the 'normal' values are indicated by symbols. (After G. S. Ljungdahl [2 a])

it is a boundary (dS being regarded as a vector having the direction of the normal to the surface element), and i is the current-density at dS. If the total current-flow through the curve is A amperes,

$$\int \boldsymbol{F} \cdot \boldsymbol{ds} = 4\pi A/10. \tag{3}$$

Since F is known only at or very near the surface, such line integrals are practically confined to the surface, and will therefore depend on any currents crossing it from earth to air or vice versa. The existence of such currents might also be deduced, at least if F were adequately known all over the earth, by an attempt to determine the surface-value of V separately from X and from Y; the two determinations should agree only if no currents cross the surface. If, however, V is thus determined from data for X and Y over less than the whole sphere, discrepancies between the two values do not necessarily indicate earth-air currents.

3.7. Earth-air currents. The existence of such currents is indicated by atmospheric-electric observations (by inference from the observed potential gradients and electrical conductivity of the air). Their magnitudes are small, being in general of the order 10^{-6} to 3×10^{-6} ampere per sq. km.; the usual direction is downwards. From (3), the average magnetic field component along the circumference of a square kilometre is found to be of the order 10^{-11} gauss $= 10^{-6} \gamma$. It would therefore be impossible at present to detect such small currents by line integrals of F, or determinations of V, as described in §6. Even if i were as much as 3 amperes per sq. km., F would be only of the order 1γ.

Such line integrals around *large* areas, however, have been computed by Bauer, and by Dyson and Furner—in the latter case from the Admiralty isomagnetic charts for 1922—and the results have sometimes been considered as suggesting the possible existence of currents 10^5 times as great as those just mentioned, i.e. of the order 10^{-1} ampere per sq. km.

By considering line integrals along parallels of latitude Bauer [5, 6] was led to conclude that there are upward vertical currents around the poles and downward currents in the tropics. Dyson and Furner [9] made similar calculations independently and obtained results in general accordance with those of Bauer. But they examined the question further by dividing the belt of latitude between $\pm 30°$ into six parts, each of $60°$ breadth in longitude; the current-densities for the six regions (as calculated from their figures) are as follows. The current

(positive, vertical, and upwards) is expressed in units of 10^{-3} ampere per sq. km.

East longitude	0°	...	60°	...	120°	...	180°	...	240°	...	300°	...	360°	Mean
Current-density	+15		−12		−8		−25		−54		+4			−13

The values for the separate regions deviate considerably from the mean, and show no apparent connexion with the distribution of land and water. For this reason Dyson and Furner appeared to doubt the reality of the currents, though they remarked that X, which enters into these calculations, is probably less well determined than Y, since the isomagnetic chart for H was constructed on a smaller scale than that for D. In the line integrals along parallels of latitude Y alone enters; its mean values round the parallels for 30° N., 10° S., and 40° S. are found by Dyson and Furner to be +0·0004, −0·0000, and −0·0004 Γ (or +40, 0, and −40 γ).

According to Angenheister [7] the current-densities in various zones of latitude are as follows, in the same units (milliamp./sq. km.) as before:

Zone	90° N.	...	50°	...	30°	...	10° N.	...	10° S.	...	30°	...	50°	...	90° S.
Current-density	+21		+19		−30		−12		+18		−16		+18		

Peters [8] has examined the question in a different way by drawing curves orthogonal to the lines of horizontal magnetic force; these should close if there are no earth-air currents. He finds that curves starting on the Greenwich meridian at latitudes 30° N. and 30° S. and drawn eastwards, always normal to H, on their return to this meridian are respectively about 53 km. S. and 112 km. N. of their starting-points. This corresponds to a systematic deviation of the declination by about 3′ east on the northern hemisphere and 9′ west on the southern hemisphere. These and other similar results, based on the same data as used by Bauer, are in general accord with Bauer's conclusions.

Agreement as to the reality or otherwise of such large earth-air currents as 0·1 ampere/sq. km. is only likely to be reached as a result of more accurate and detailed measurements of D and H than are at present available. According to Schuster [38] it should be possible to reach a decision by extremely careful measurements at short intervals round a contour of small circumference (about 4 km. square). Until such measurements are made, it is advisable to regard the hypothetical earth-air currents as probably due merely to lack of accuracy in the magnetic data (see Schmidt [10 b]).

3.8. Secular variation; isoporic charts. While it is sufficiently difficult to visualize the distribution of the average magnetic field, it

is even more difficult to form a pictorial conception of the way in which the earth's field F is *changing*, either by secular or transient variation. The rate of the secular variation in each element at any epoch (reckoned as the change per year, positive when the change is an increase) can, however, be represented by *isovariational* charts, or *isoporic* charts, a term proposed by Harradon (derived from the Greek word for movement, and used by Fisk [11]). The Admiralty world magnetic charts include, in addition to the large-scale isomagnetic chart, a smaller isoporic chart for each element D, H, and I.

Figs. 13–17 show isoporic charts constructed by Fisk [11], who with Fleming emphasized the following significant facts about the isoporic curves. (1) Isopors tend to form closed ovals around certain foci of rapid annual change. (2) In general the accelerations or changes of rate from year to year are large near these foci. (3) The areas of rapid change are not permanent but may appear, or undergo radical changes in form or position, in so brief a time as one or two decades. (4) These foci are at present practically all in one hemisphere, namely the one bounded by the meridians 90° W. to 90° E., and containing most of the land of both hemispheres. The apparent absence of any important focus in the Pacific Ocean is not due to the scarcity of observations there; sufficient have been made to reveal the presence of one if it existed, though they might be inadequate to enable its outline or behaviour to be traced.

The isoporic charts are drawn on an equal-area base-map, which better represents the relative sizes of the areas of special interest than does the Mercator projection, and permits the extension of the lines across the polar regions. Fig. 13 for declination shows four well-defined foci and several of moderate activity. The lines surrounding the South American focus have expanded from the time of the earliest magnetic observations made there, about 1904, until about 1917, and have since been slowly contracting. The South African and Indian Ocean foci present an interesting condensation of isopors near Madagascar: declinations at these two centres are changing relative to one another by nearly half a degree a year. The South African focus appears to be contracting and possibly moving south-eastwards. The European focus seems to be of recent growth, since the rates of annual change over most of Europe in 1905 were between 2·5' and 5·0', and nowhere approached the subsequent maximum of 13'. The zero-isopor crossed central Siberia in 1905, and has since drifted slowly westwards to near the Ural mountains. During the same period the lines round the
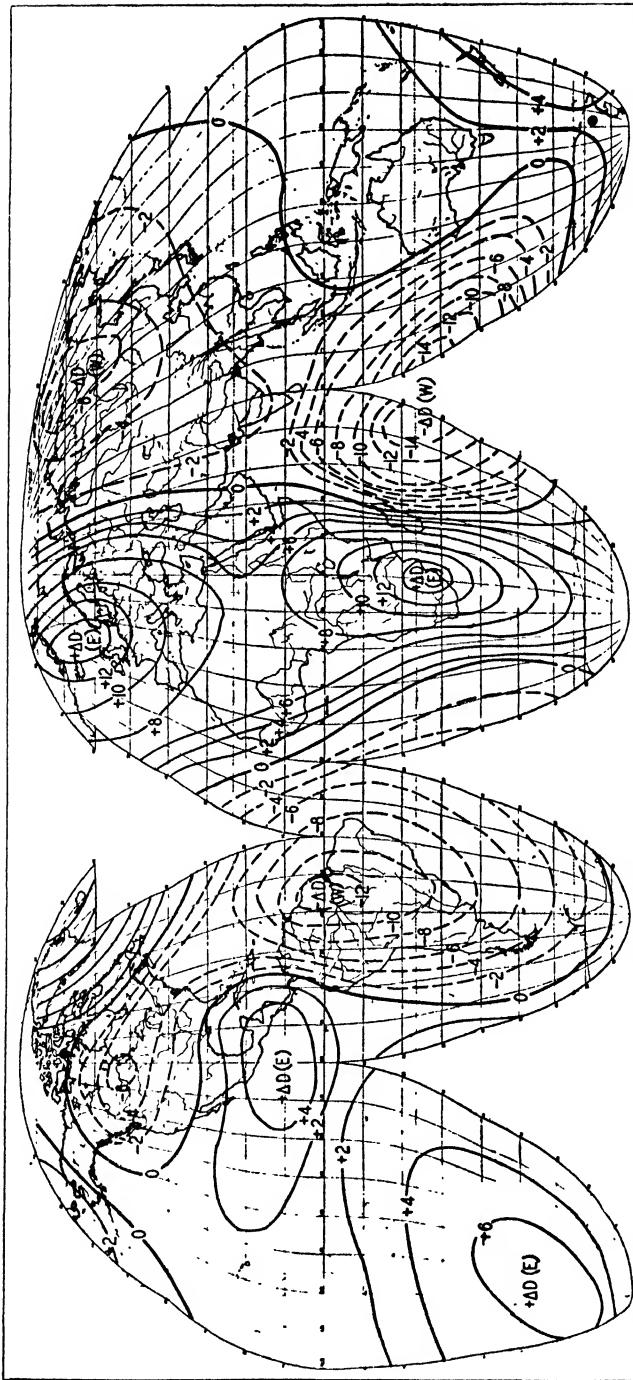
FIG. 13. Isoporic chart for the magnetic declination D, showing the lines of equal annual change of D (in minutes per year) for the approximate epoch 1922. The isopors in high latitudes, especially near the magnetic poles, are uncertain. (After H. W. Fisk [11])
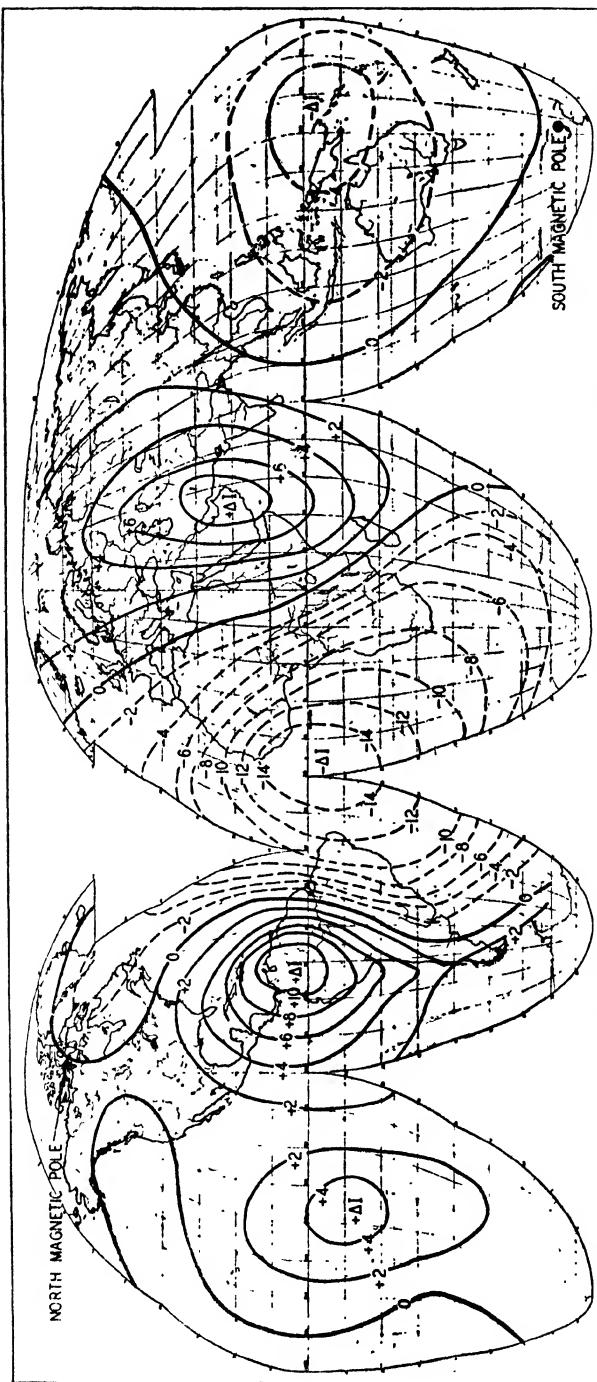
FIG. 14. Isoporic chart for the magnetic inclination I, showing the lines of equal annual change of I (in minutes per year) for the approximate epoch 1922. (After H. W. Fisk [11])
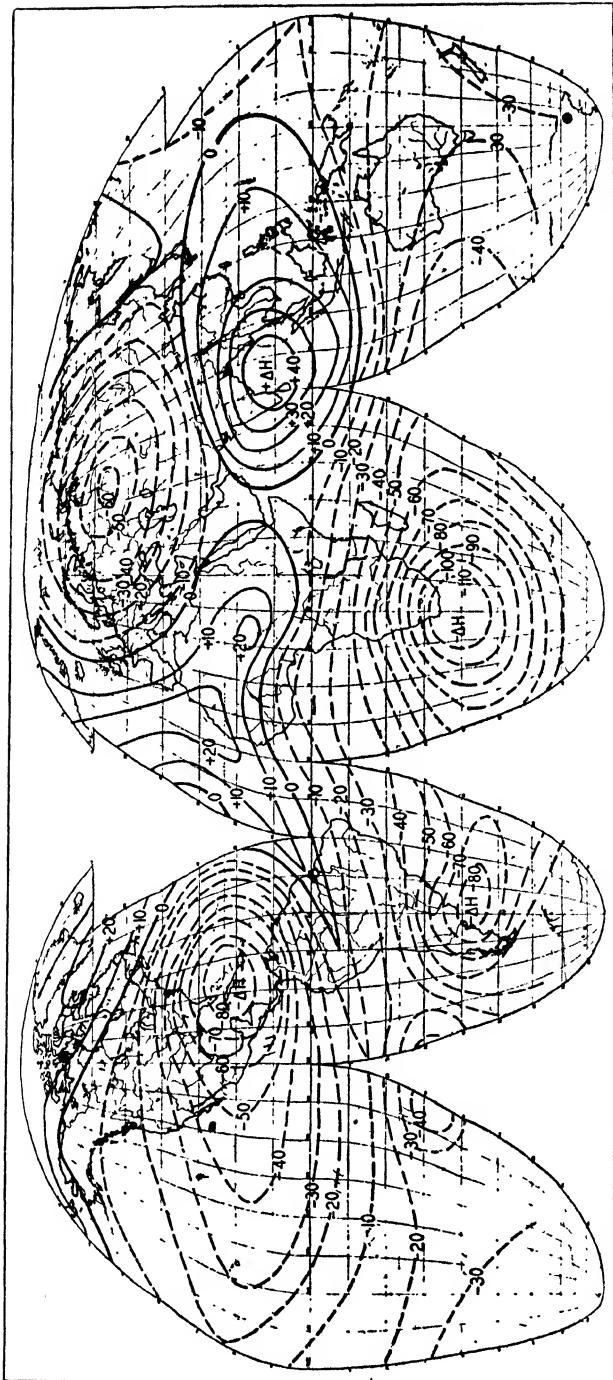
FIG. 15. Isoporic chart for the horizontal magnetic intensity H, showing the lines of equal annual change of H (in gammas per year) for the approximate epoch 1922. (After H. W. Fisk [11])
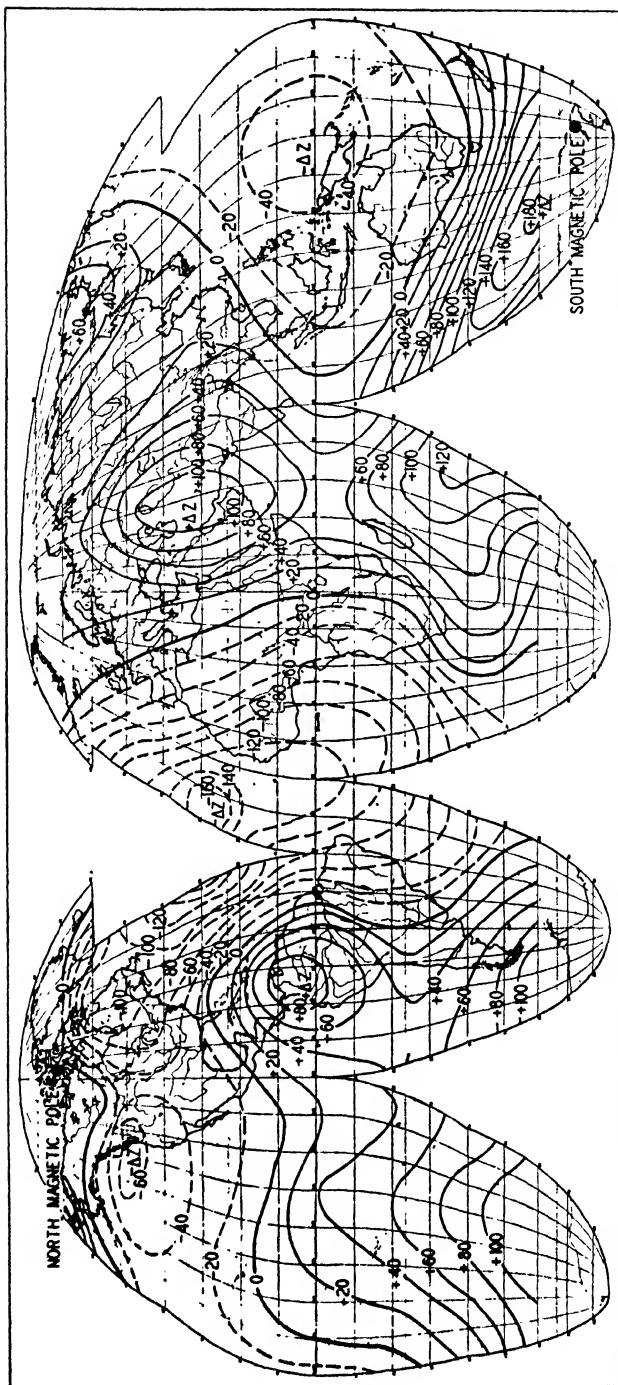
Fig. 16. Isoporic chart for the vertical intensity Z, showing the lines of equal annual change of Z (in gammas per year) for the approximate epoch 1922. (After H. W. Fisk [11])

(119)
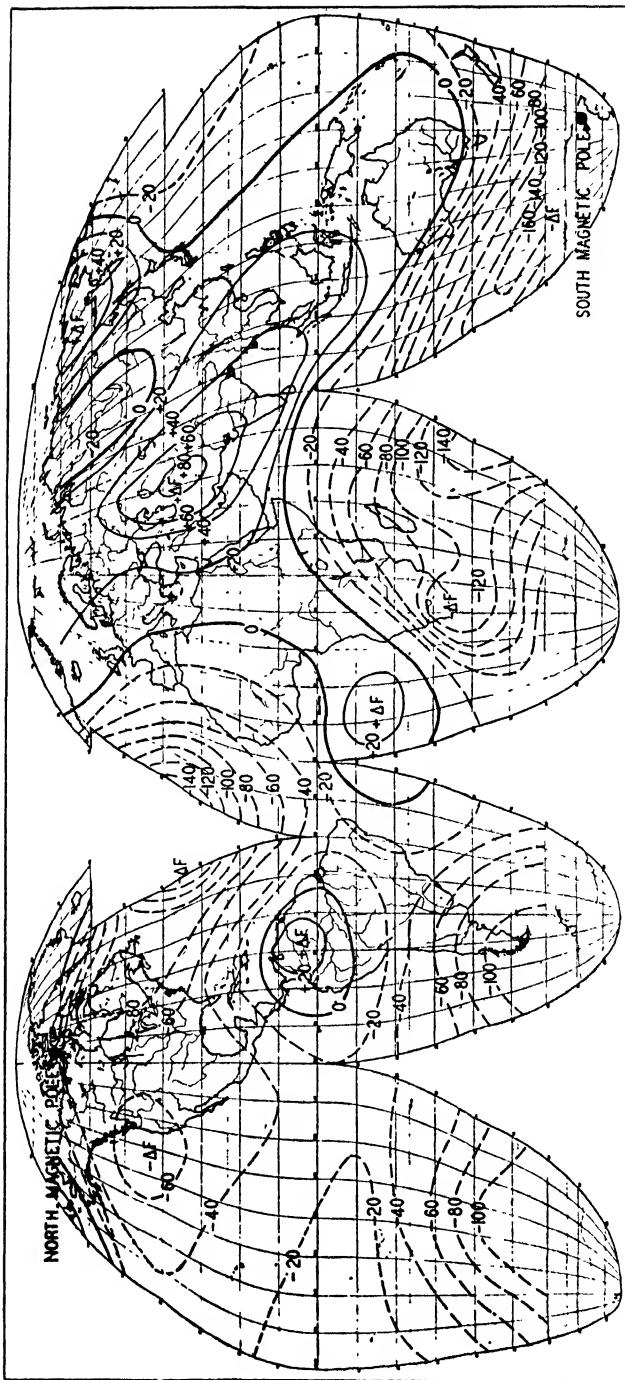
FIG. 17. Isoporic chart for the total magnetic intensity F, showing the lines of equal annual change of F (in gammas per year) for the approximate epoch 1922. (After H. W. Fisk [11])

North Sea focus expanded; those moving eastwards crowded together in eastern Russia. At places between these regions of eastward and westward movement the rate has remained nearly constant, as at Kazan.

The inclination isopors in Fig. 14 show three principal foci, one in western South America, in which the dip is increasing (algebraically), another of less intensity in Iran, and a very intense focus of decreasing dip in the tropical Atlantic. These isopors also have shown very decided movements even during the comparatively short time for which observations are available. Before 1917 annual rates as high as $+16'$ or $+18'$ occurred in Colombia and Ecuador, which have since fallen to $+10'$ or $+12'$. While the ovals about this focus have contracted, those about the Atlantic focus seem to have expanded. One joint effect of this is a westward movement of the zero isopor; for example, the point at which it crosses the Amazon has advanced considerably up that river.

As regards intensity, the horizontal component (Fig. 15) is decreasing in the southern Argentine, in the Caribbean Sea, in South Africa, and in north-west Siberia, whereas there is only one well-marked and relatively small region of increasing horizontal force, with its focus in southern India. South American observations before 1917 indicated increasing rates over most of the Amazon basin, but have later shown small decreasing rates. In the Caribbean the rate of decrease has greatly diminished in recent years, while there is a slight indication that rates have increased about the Patagonian centre. The high decreasing rates in South Africa are notable chiefly because they represent a relatively high percentage of the total horizontal component—in recent years the annual decrease has been one part in 135: this high rate is known to have persisted for a long time, the decrease in thirty years being about $3,000\gamma$, nearly a fifth of the present value. The latest observations indicate that the rate of decrease has now begun to diminish. The area of increasing horizontal component lies in the northern hemisphere, except for small areas in the Indian Ocean and near the mouth of the Amazon. The maximum annual increase, about 40 to 50γ, is less than half that found at several of the negative centres. In the aggregate, therefore, the earth's horizontal field is decreasing at present.

The vertical-intensity isopors (Fig. 16) resemble those for inclination, the chief exception being that the Atlantic focus is more to the north. The apparent focus of increasing vertical force south-west of Australia is doubtful.

The total-intensity isopors (Fig. 17) show that the area of increase is much less than that of decrease, and that the maximum rates of increase are also less than those of decrease. Here again the area of decreasing intensity south-west of Australia may be shown too intense.

The total change from 1885 to 1922—that is, between the epochs chosen for the analysis of the earth's field by Schmidt (1885) and by Bauer, Dyson, and Furner (1922)—is shown by the two charts Figs. 18 and 19. Although some of these differences are certainly due to our incomplete knowledge of the field in 1885, the general character and magnitude of the changes can be trusted. The greatest horizontal vectors for the differences reach $4,000\gamma$, and the greatest change in Z is as large as $6,000\gamma$.

Heck [12] has published an interesting series of charts for the D and H isopors in North America since 1785, of which a selection is given in Figs. 20-3. The change is striking, and can be described roughly by the shift in the position of the lines for zero change in D (Fig. 21). These lines lay in the north-to-south direction from 1785 to 1900; in 1785 the line lay near the meridian 72° W., then it shifted westwards till 1875, when it reached the meridian 118° W. After that, it returned eastwards till 1900, when it lay near 100° W. From that date it began to turn, attaining the direction north-west to south-east in 1915, till it became approximately east to west in 1920 (along 30° latitude) and 1930. The H isopors show how the decrease has accelerated.

Reich [20] and Schwinner [32] have studied the Z isopors in Central Europe in relation to geology; Schmidt [19 a] gave a quite different interpretation.

The *stereogram*, Fig. 24, shows the secular variation of the magnetic field vector at the intersections of the meridians 0°, 20°, etc., with the circles of latitude 60° N., 40° N., ..., 40° S., as derived, for the epoch 1925, from Fisk's isoporic charts for D, H, and I, and expressed as annual changes ΔX, ΔY, and ΔZ. The vectors with the components ΔX, ΔY, ΔZ are represented in the stereogram as pointed cones. For instance, off the north-east coast of South America, in longitude 40° W. on the equator, $\Delta X = -21\gamma$, $\Delta Y = -72\gamma$, $\Delta Z = -99\gamma$, and the vector therefore points towards south, west, and upwards. The scale can be judged from the distance between successive meridians (15° apart) on the chart; this length corresponds, on the scale used for the magnetic vectors, to 61γ; it is also shown by the lines drawn near the corners of the charts, which appear perpendicular if viewed through a stereoscope, and which correspond in length to 100γ downwards
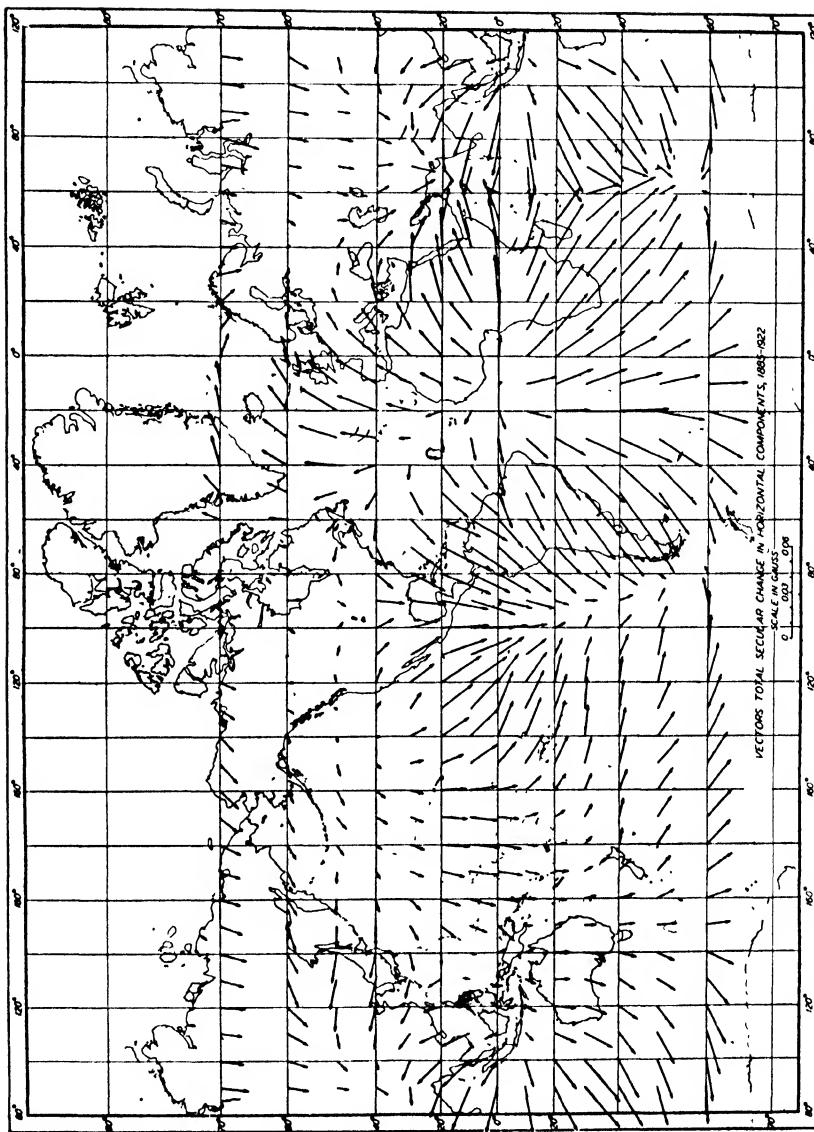
Fig. 18. Vectors showing at different points of the earth the total change of the horizontal component of the geomagnetic field from 1885 to 1922. (After A. G. McNish)

(123)
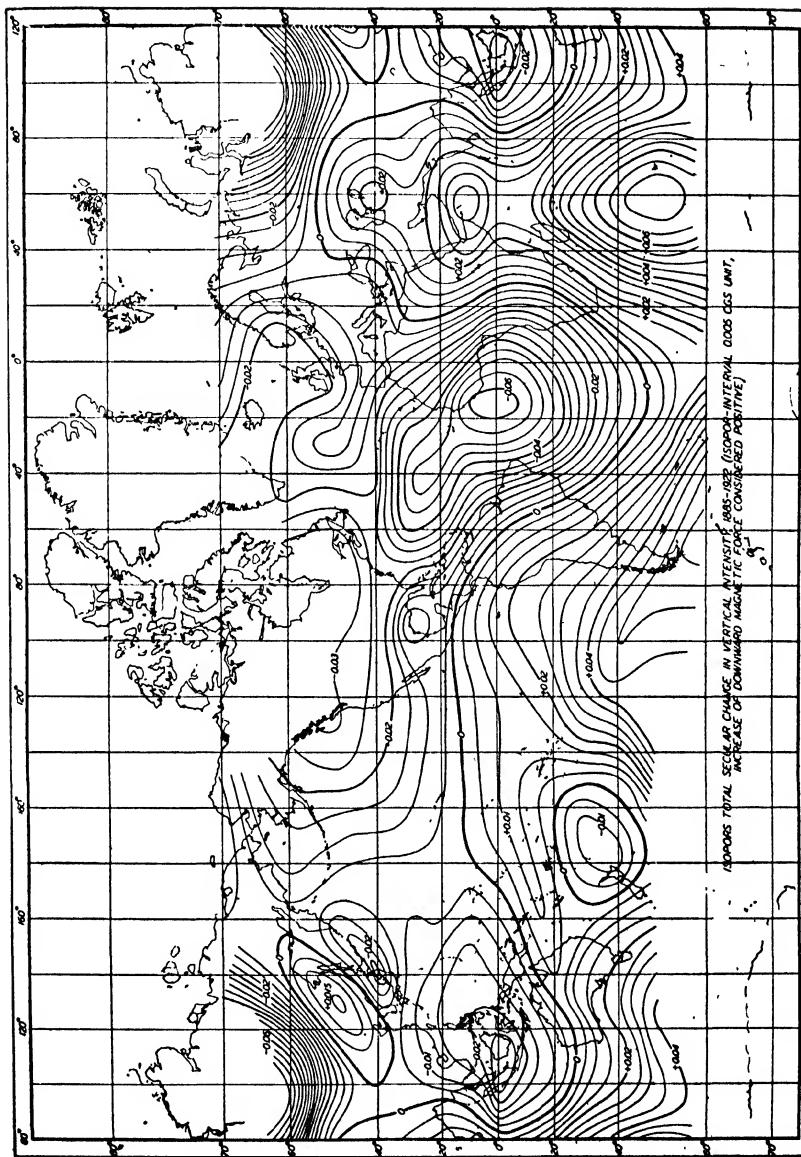
FIG. 19. Isopors showing the lines of equal total change of the vertical magnetic intensity Z from 1885 to 1922; the isopors are drawn at 500 γ intervals. (After A. G. McNish.)
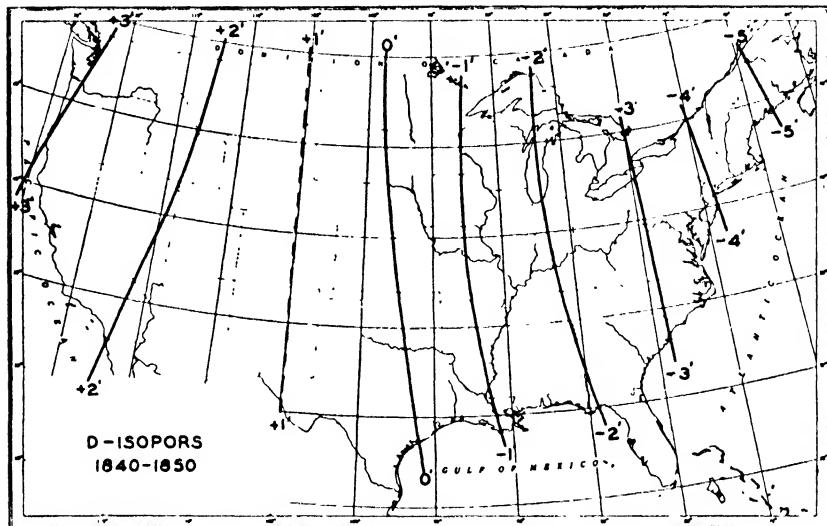
Fig. 20 *a*

Fig. 20 (*a-e*). *D* isopors for part of North America, showing the lines of equal annual change of magnetic declination *D* (in minutes per year), during alternate decades from 1840 to 1930. The positions of the zero isopor at these and other epochs are shown in Fig. 21. (After N. H. Heck)
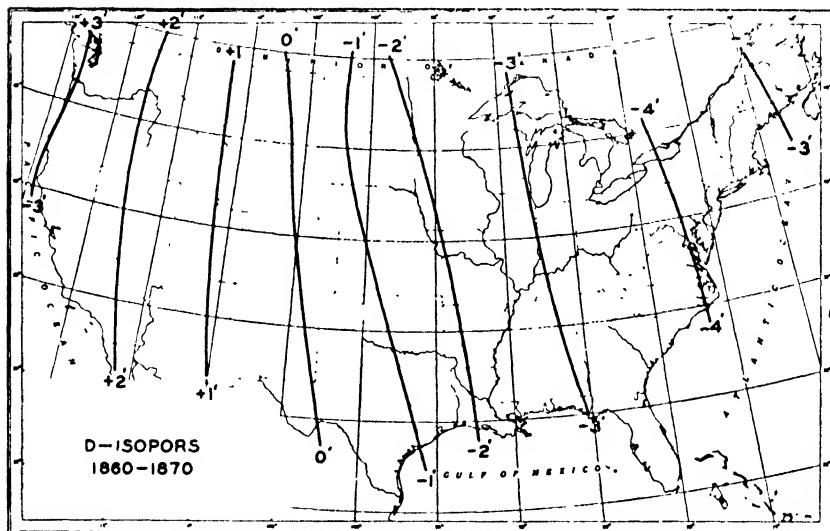
Fig. 20 *b*
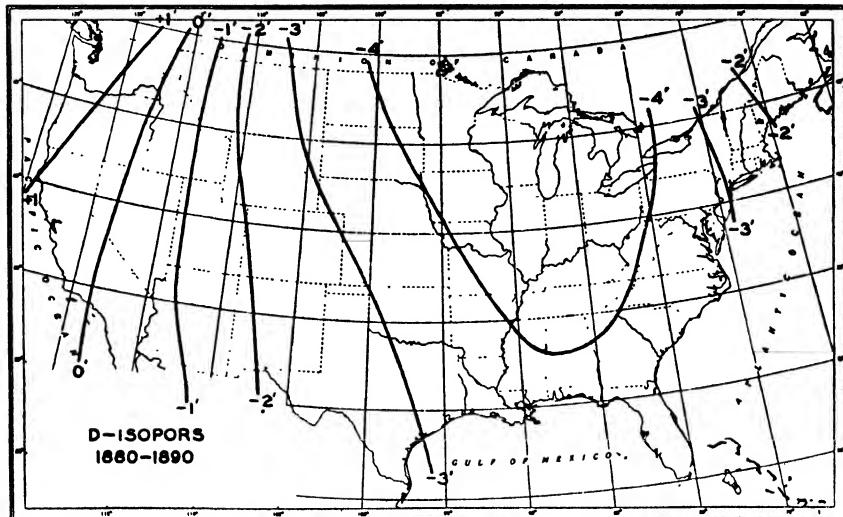
FIG. 20 *c*. See p. 124
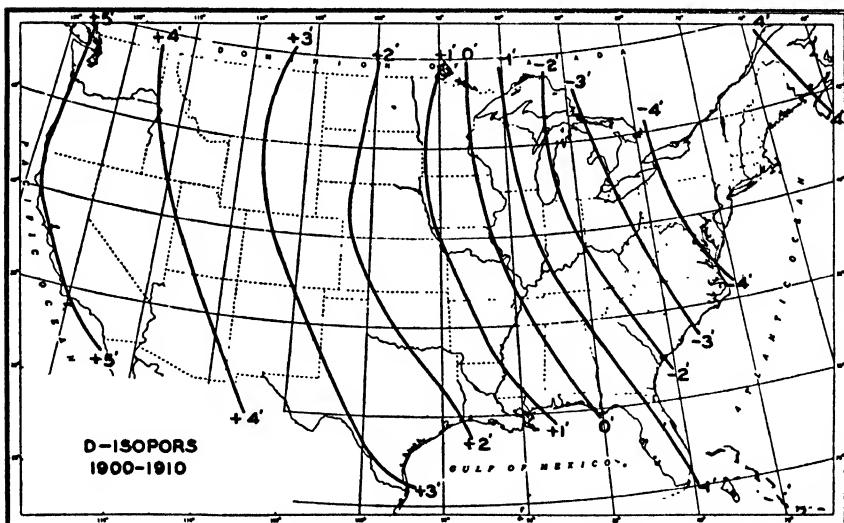
FIG. 20 *d*. See p. 124
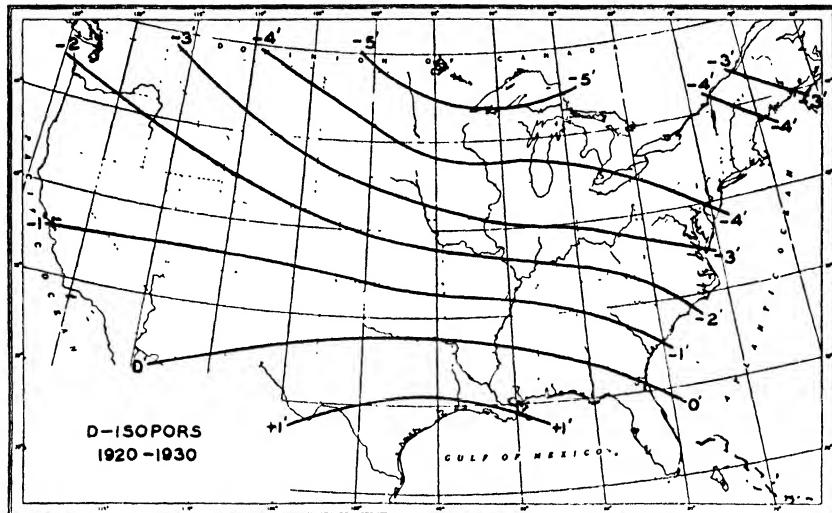
FIG. 20 *e*. See p. 124
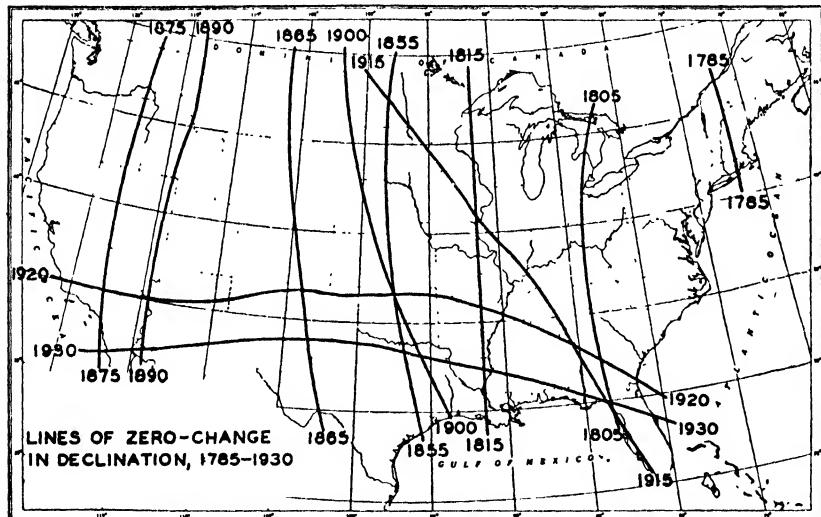
FIG. 21. Positions of the *D* zero isopor, or line of no change of declination, over part of North America, at various epochs from 1785 to 1930. (After N. H. Heck)
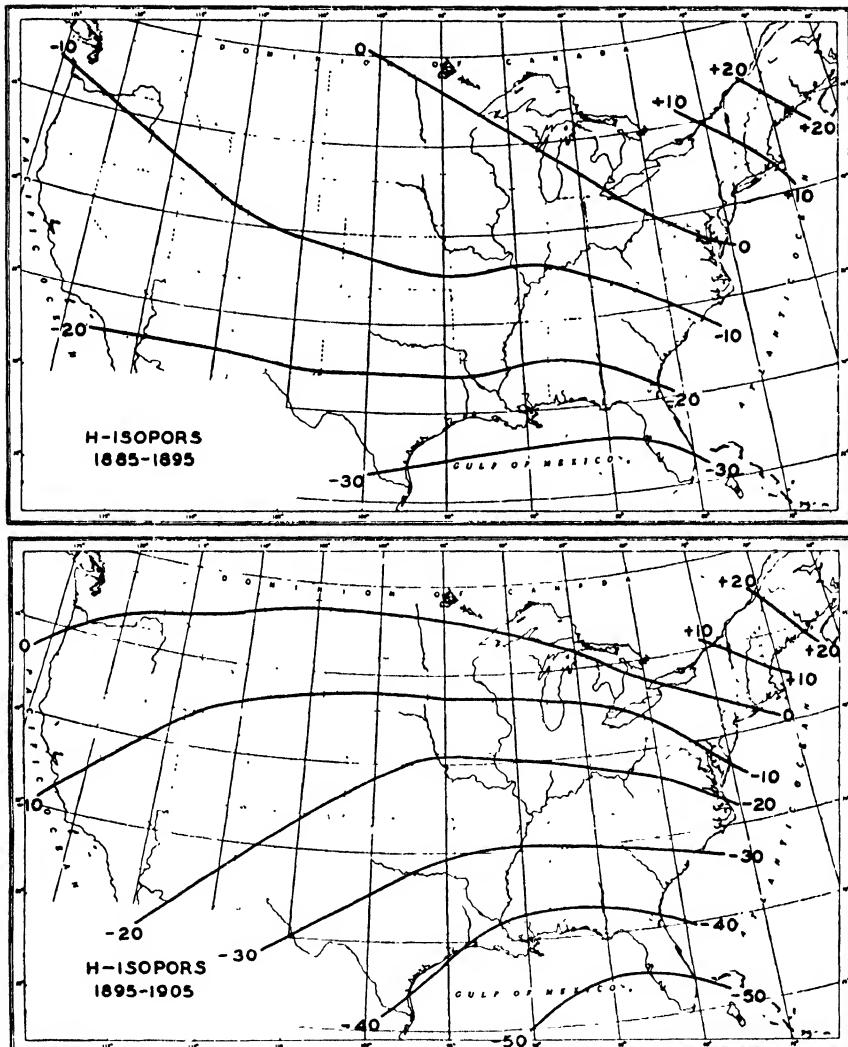
FIG. 22. *H* isopors for part of North America, showing the lines of equal annual
change of the horizontal magnetic intensity *H* (in gammas per year) for two
successive decades. (After N. H. Heck)
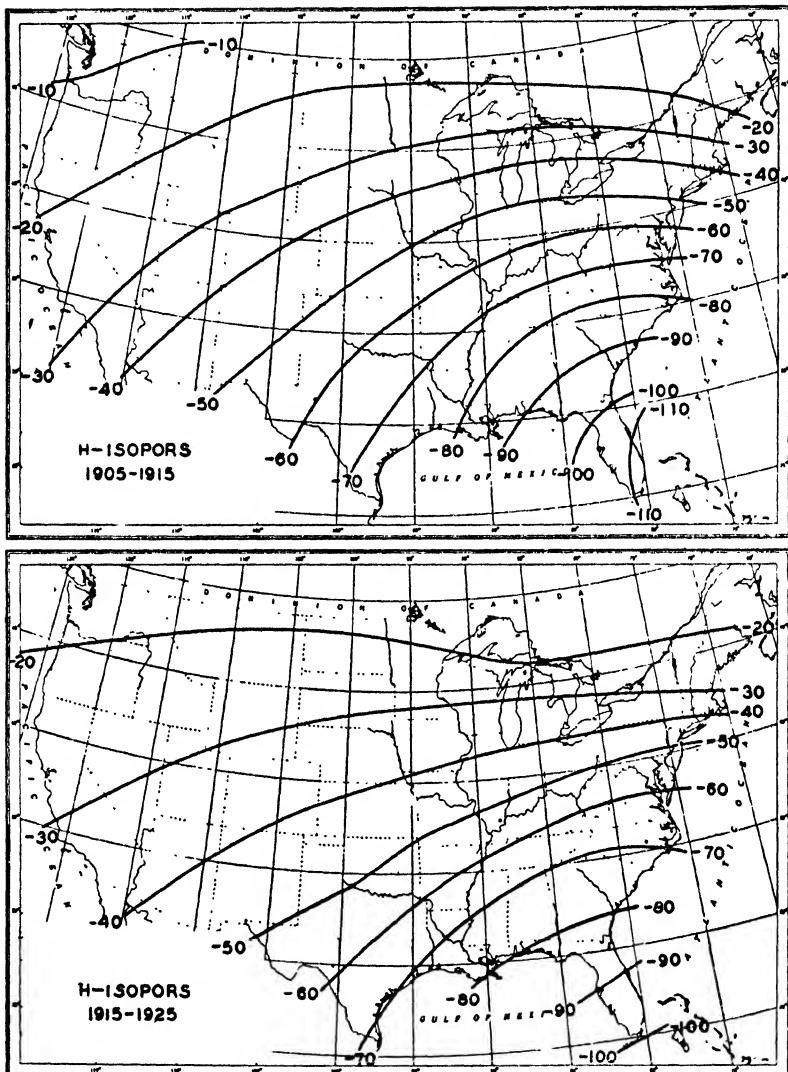
Fig. 23. *H* isopors, as in Fig. 22, for two further decades

and 100γ upwards or 200γ in all. The longest vector on the chart is found in 40° N. and 40° W., where $\Delta X = +38\gamma$, $\Delta Y = +10\gamma$, $\Delta Z = -159\gamma$.
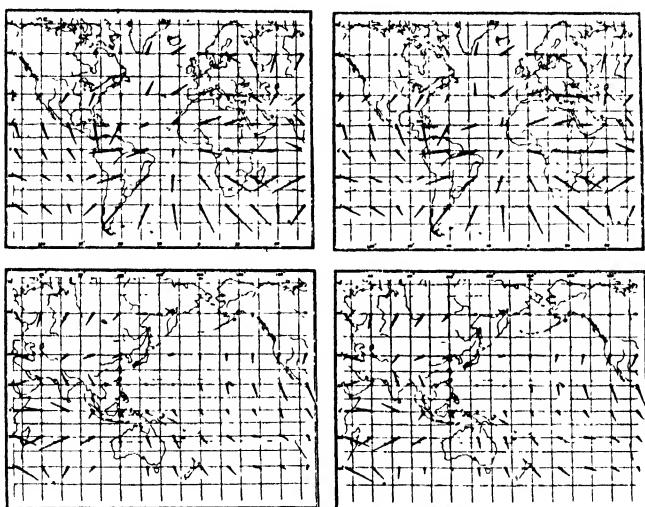
FIG. 24. Stereograms showing the vectors of the annual secular magnetic variation for 1925; above, Europe, Africa, America; below, Asia, Australia, and the Pacific Ocean

3.9. Long series of secular variation observations. Our knowledge of the secular variation is incomplete, because only at a few stations in Europe, such as Rome, London, and Paris, do the data extend over so much as three or four centuries. Moreover, the older observations refer solely to the direction of the magnetic force (D and I): measurements of intensity on an absolute scale are available for barely a century. Fig. 25 indicates the variation of the direction of F at London during nearly four centuries, by the path of the trace of F (drawn from a fixed origin about 20 cm. above the paper) on a plane normal to the mean direction of the force. The vector is seen to have completed about three-quarters of a conical surface, having a mean angular radius of about 6°; the range in I is about 8°, and in D is from 30° to 35°. If the surface continues to be described as the past course suggests, it appears likely to close (or nearly so) after a total period of about 480 years. Other European stations show similar changes, but with turning-points in D and I at dates rather different from those for London; the curves for Boston and Capetown are more elongated in the meridian plane. The rate of description of the cone shows some

variations, but the comparative uniformity of the rate during the past three centuries is the more noteworthy feature. The variation of intensity shows similar properties, which may be illustrated by the following
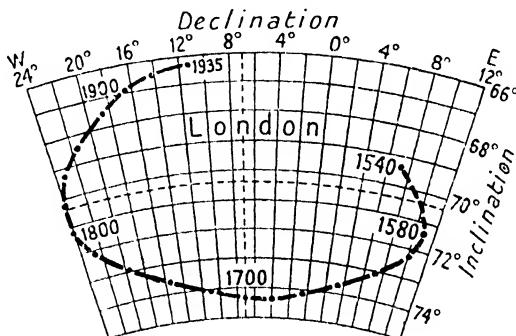
FIG. 25. The direction of the magnetic force at London since A.D. 1580.
(After L. A. Bauer)

mean values of H at Greenwich from 1853 to 1933; the changes ΔH are given in the unit 0·0001 gauss (or 10γ) per decade, or γ per year.

1853	$H = 0·1734$ gauss, $\Delta H =$		1903	$H = 0·1852$ gauss, $\Delta H = +·0021$
1863	0·1764	+·0030	1913	0·1853	+·0001
1873	0·1793	+·0029	1923	0·1843	−·0010
1883	0·1812	+·0019	1933	0·1835	−·0008
1893	0·1831	+·0019		

Fig. 26 shows the annual means for all elements at Potsdam, 1890–1935.

Attempts have been made to study the past secular variation of the earth's field as a whole, from the sixteenth century, but the results obtained are subject to doubt. According to van Bemmelen [18, 19] the N. dip-pole has moved southwards since 1600. Carlheim-Gyllensköld [18.26, 18.27] inferred that the various spherical harmonic constituents of the field either remain constant, or vary periodically in magnitude, while their axes revolve round the earth's geographical axis, with periods of 3147, 1381, and 454 years for the constituents P_1^1, P_2^1, and P_2^2 respectively; but a subsequent investigation [20 c] renders such conclusions improbable (see 18.11).

3.10. Secular variation as a regional phenomenon. Bartels [20 c] made a spherical harmonic analysis (which will be discussed in 18.11) of the secular variation for the interval 1902–20, using only reliable data derived from fourteen widely distributed magnetic observatories (Fig. 27). It appears from this analysis that the secular variation is not a planetary or world-wide phenomenon like the main field, in which the F_1 field is so outstanding.

It seems unlikely that the present decrease of the earth's magnetic moment, about 1/1000 of the whole per year, has been proceeding at

ANNUAL MEANS, MAGNETIC ELEMENTS AND FORCE
COMPONENTS, POTSDAM, GERMANY, 1890-1935
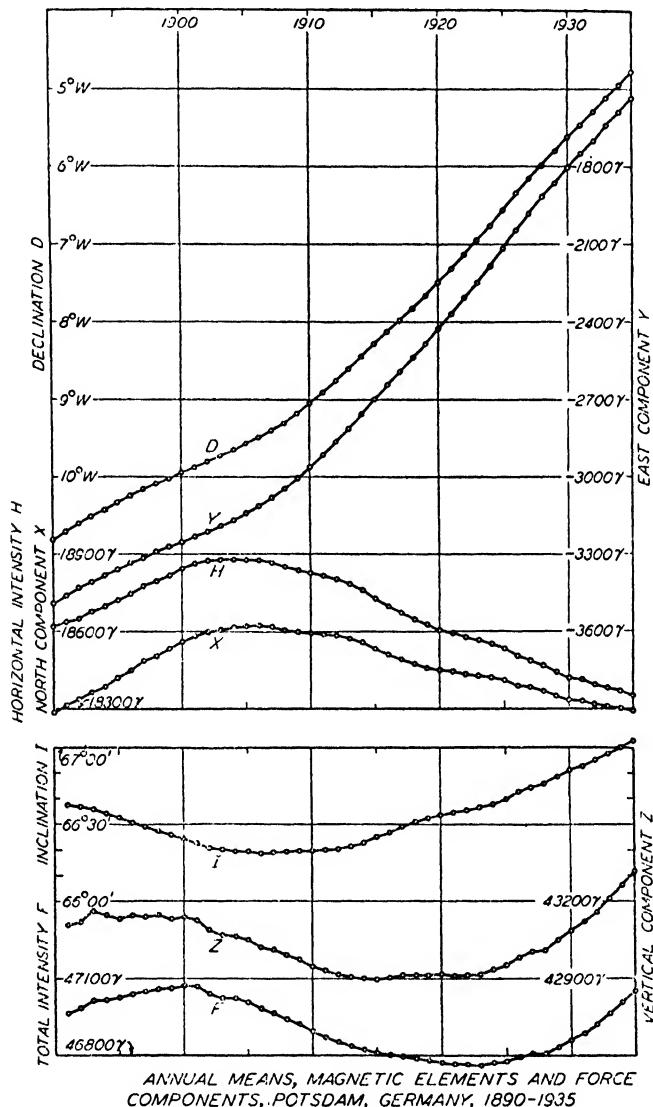
FIG. 26

the present rate for centuries, because this would imply such a high intensity of the field some thousands of years ago.

The secular variation data for six subdivisions of the interval 1902–20 indicate the regularity of the progression of the secular variation in time.

Although the secular variation is found to be a regional rather than a world-wide phenomenon, its course over large areas such as Europe (1/50 of the earth's surface), for a period of several years, can be approximately represented as a linear function of latitude and longitude. Even at locally disturbed stations, the secular variation of the
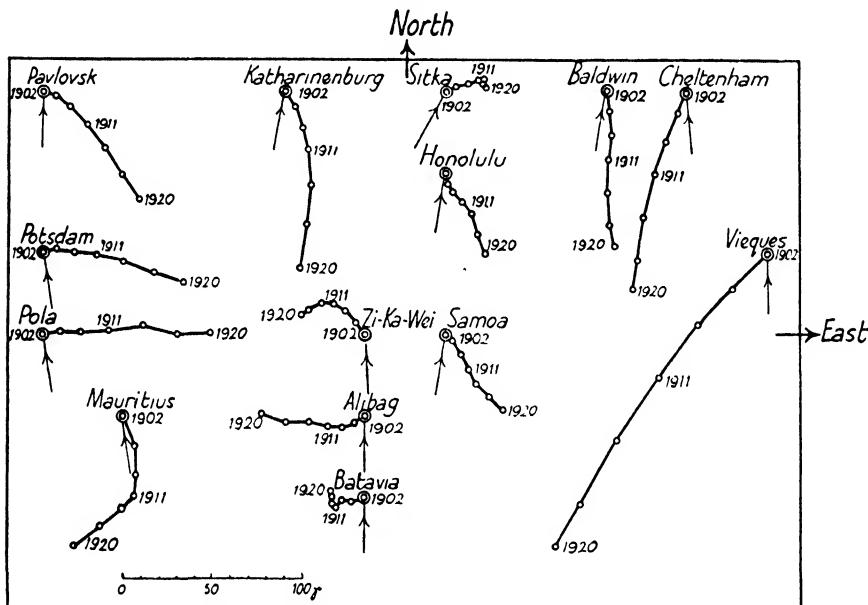
FIG. 27. Vectors showing for fourteen magnetic observatories the change in the horizontal component of the geomagnetic field from 1902 to 1920

force components X, Y, Z is practically the same as that found in the undisturbed surroundings. For example, for Central Europe Schmidt [13] found the following formula for the secular variation at the epochs 1895, 1905, and 1912, expressed as an annual change; the point of reference is in latitude 50° N., longitude 10° E., and x is the latitude difference (i.e. the latitude in degrees, minus 50°), and y the longitude difference (longitude in degrees, minus 10°).

Declination	1895	$5 \cdot 2$	$+0 \cdot 07x$ $-0 \cdot 03y$
(minutes of arc)	1905	$5 \cdot 1$	$-0 \cdot 18x$ $-0 \cdot 05y$
	1915	$8 \cdot 9$	$+0 \cdot 04x$ $+0 \cdot 05y$
Horizontal intensity	1895	$+24$	$-0 \cdot 6x$ $-0 \cdot 4y$
(unit γ)	1905	-3	$-0 \cdot 2x$ $-0 \cdot 7y$
	1915	-12	$-1 \cdot 3x$ $-0 \cdot 6y$

For the vertical intensity Z, however, Reich [20] and Schmidt [19a] found a non-linear distribution over Germany.

3.11. Effect of the sunspot cycle on the secular variation. Normal values and secular values.

Transient magnetic disturbance, on the average, lowers the value of the horizontal intensity (Ch. VI). The 11-year cycle in the frequency of magnetic disturbance, similar to the sunspot cycle (Ch. XI), therefore causes the values of H, near the sunspot-maximum, to be systematically lower than near sunspot-minimum; e.g. at Potsdam the difference is about 30γ. This difference affects the rate of annual change in H; in years of increasing sunspot numbers and increasing magnetic activity, this rate will be algebraically less than in years of decreasing sunspot numbers. Fisk [20 d] demonstrated this effect in the observations of H at ten stations, for the period 1903–27 (Fig. 28). The broken lines give the observed changes ΔH from year to year; the solid lines represent ideal curves of a simple form (the ordinates being equal to a constant plus three cosine terms of 48, 24, and 16 years' period) fitted to the observations. The relative diminution of the observed ΔH in the years preceding the sunspot-maximum years 1917 and 1928 is evident. The declination and vertical intensity are also affected by the 11-year cycle, but to a less extent.

In order to eliminate this effect, Schmidt [G 51] has proposed that an average secular variation should be computed from so-called 'secular values' of the magnetic elements, these being averages for 10 years (which are practically the same as 11-year averages). His 'normal values' (*Normalwerte*) are average values for 12 months, and are ascribed to the middle of that time-interval.

3.12. Indirect inferences concerning the secular variation.

Apart from the actual observations of declination, going back to about the year 1500 (the older ones have been collected by van Bemmelen [18, 19]), there is the possibility that the orientation of old churches and other buildings may have been settled, when they were being planned, by means of a magnetic compass. If so, the azimuth of their central lines would indicate the declination at the time of building, and changes in the azimuths between earlier and later parts, as, for instance, in the Dom (cathedral) at Magdeburg, Germany, might be explained by the secular magnetic variation. Nippoldt [24], while reviewing papers by Wehner [23] on German churches, found that several churches seem to give the declination at the time of their foundation, within an error of 2 degrees, and advocated a more detailed study of the question.

Folgheraiter [21], Mercanton [27, 28], Koenigsberger [22], and others have studied the direction of magnetization in old pottery, rocks, and
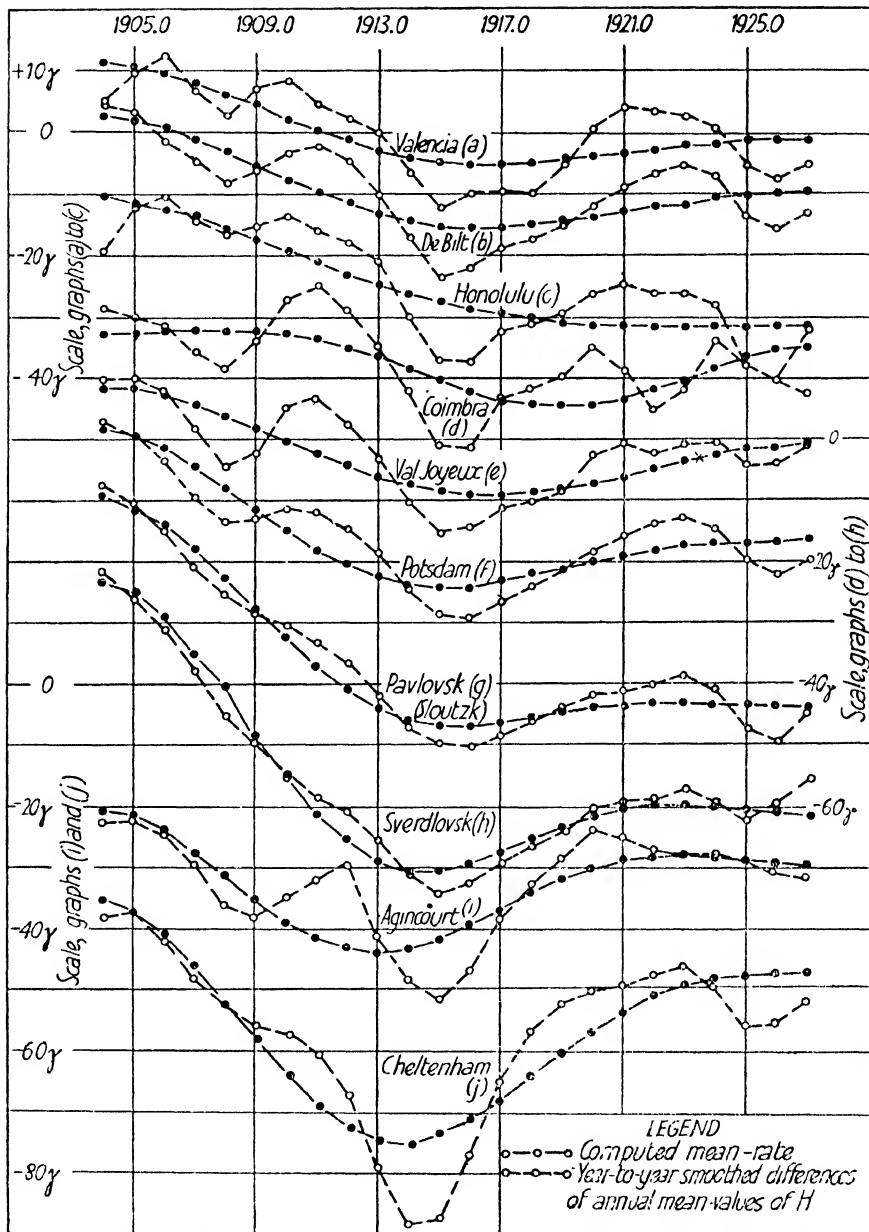
FIG. 28. The change from year to year in the horizontal magnetic intensity, 1903–27, at ten observatories. Broken lines: the differences between successive annual mean values of *H*, slightly smoothed. Full lines: the differences computed from simple formulae which approximately fit the broken lines. The difference between the two lines for each observatory indicates the influence of the sunspot cycle. (After H. W. Fisk [20 d])

lava fields. Just as a piece of hot steel, when cooled by being put into cold water, becomes magnetic under the inducing influence of the earth's magnetic field, so also it appears that streams of lava acquire a magnetic moment at the time of cooling, and it seems possible that they retain it without any appreciable effect from later changes in the earth's field. Chevallier [29, 30], in extensive studies, has examined many samples of lava from Mt. Etna (Sicily). The intensity of magnetization is least in the inner parts of the streams, where it may be as little as 0·008 c.g.s.; at the surface it is often 0·015, and may even be as high as 0·1. Different pieces of lava from the same stream may show different directions of magnetization, but a number of samples from different places in the lava stream give an average direction agreeing with that of the earth's field at the time when the lava cooled.

Vases and bricks of baked clay [26] have a high coercive force, according to experiments by Koenigsberger [22] and others, its value sometimes approaching the value for pure magnetite: and, as in the case of rocks, it is unlikely that they lose their remanent moment even during thousands of years. It is therefore interesting (although hardly conclusive) to find that certain Etruscan and Grecian vases of date 700 or 600 B.C., according to Folgheraiter [21], would indicate a reversal of the dip as compared with its present value.

The strongest evidence for a nearly complete reversal of the magnetic field in geological times has been provided by certain highly magnetic eruptive rocks from the Transvaal (South Africa); these palaeozoic volcanic dikes, formed about 100 million years ago, show a direction of magnetization definitely opposite to the present field, according to Gelletich [31], whose work is discussed more fully in 4.12.

Evidently such objects as lava, vases, bricks, or eruptive rocks can provide indications as to the former state of the geomagnetic field only if it can be made reasonably certain that their position as well as their magnetic state has remained unchanged since they were magnetized.

Koenigsberger [22 b] has recently given an extensive bibliography of studies on the residual magnetism of eruptive rocks. Thellier [26] has developed convenient and accurate apparatus for such studies, and has made extensive measurements on rocks, bricks, and other objects of baked clay. After reviewing the evidence afforded by his own and other measurements, he concludes that the permanent magnetization of rocks is ill-defined, and gives no safe basis on which conclusions as to the past state of the earth's magnetism can be arrived at. Objects of baked clay, on the contrary, appear well suited to this purpose, when

adequate particulars concerning them are available. By measurements on the magnetism of French bricks of known age, dating from A.D. 1400 onwards, he has obtained what he regards as a reasonably probable curve showing the variation of magnetic dip from this time until actual observations of dip were made, at about A.D. 1780, for Paris. The later part of his inferred curve fits on well with the actual observations made at Paris, and also runs reasonably parallel with the earlier part of the dip-curve for London (Fig. 25). Thellier has also proposed a method of determining the past *intensity* of the earth's field, by experiments on old bricks, and this method seems to promise valuable and interesting results.

MAGNETISM AND GEOLOGY. MAGNETIC PROSPECTING

4.1. The purpose of magnetic prospecting. Deposits of iron ore or other magnetic minerals generally distort or 'disturb' the otherwise even trend of the isomagnetic lines, partly by the field of their own permanent magnetization, and partly by the additional magnetism induced in them by the general field of the earth. When the deposits are near the surface, the principal features of the resulting local magnetic anomaly or disturbance can readily be determined by means of simple instruments, such as the compass and dip-needle. This has long been done in Sweden and elsewhere. These measurements are of value in judging the form and position of the ore-deposit, and give a clue as to where to bore.

For some years past practical geology has made more and more use of the methods of applied geophysics, and magnetic prospecting is now undertaken in cases where the juxtaposed geological layers differ only slightly in their magnetic properties. This has prompted the design of specially sensitive instruments called *local variometers*, and special techniques have been developed; for the study of these the reader is referred to more extensive and special treatises [1–12] and papers.

4.2. Local variometers. The apparatus used for absolute observations in ordinary land surveys may naturally also be used in prospecting work, but this would be inconvenient, because in the detailed investigation of local anomalies it is often necessary to make observations at many stations only a few metres apart. In order to detect every important irregularity of the field and yet finish the survey in a reasonable time, local variometers of special design are required. They must be easily and safely transportable, usable everywhere without tiresome adjustments, and the quantity that is to be measured must be given by a simple reading, with only slight corrections. Since the absolute values are not of primary interest, local variometers are relative instruments, giving the departures from the normal values at one or more base stations in the vicinity.

If the observations are to be confined to a single magnetic element, it is advisable to measure the vertical intensity Z, because, as will be seen, its deviations from the normal are rather easily interpreted in regions where the inclination I is not too small. Sometimes also the horizontal intensity is measured. Usually the declination is not

measured, since it is inconvenient to procure the necessary astronomical or geodetic data which define true north.

4.3. Schmidt's field balances. The vertical-field balance devised by Schmidt [13] is a local variometer for Z based on the principle of Lloyd's balance (p. 53); it is shown in Plate 13. The needle (Plate 14 a) consists of two parallel magnetized steel blades fixed to an aluminium cube. During transport it is clamped with the knife-edge lifted off the agate-block.

In use, the instrument is put on a wooden tripod and is turned about a vertical axis so that the needle stands normal (east–west) to the magnetic meridian, as indicated by a compass. The needle, when set free, will be tilted more or less from its normally horizontal position, according to the value of Z at the station. The tilt is read by means of a telescope with scale, reflected by a horizontal mirror fixed to the needle. If n and n_0 are the scale-readings at the field station and at the base station, t and t_0 the corresponding temperatures inside the balance, the difference in Z between the two stations is computed from the formula

$$\Delta Z = Z - Z_0 = k(n - n_0) + \mu(t - t_0). \tag{1}$$

The scale-value k is generally 20 to 30 γ per division of the scale, except for prospecting in magnetic iron-ore regions, where it may be 200 γ or more. It is determined, as in the case of observatory variometers, by applying artificial fields of known strength by Helmholtz-Gaugain coils (Plate 14 b). The temperature coefficient μ is found by readings in heated rooms. In the older models of the instrument, μ was of the order of 10 to 20 γ per degree centigrade; but new compensated magnet systems (Heiland [14, 15]) have lowered the temperature coefficient considerably. The balance can also be provided with an attachment for photographic recording (Plate 14 c) or photo-electric recording [13 c].

The *horizontal-field balance*, giving ΔH, contains a vertical needle, free to swing in the magnetic meridian about a horizontal axis. Here the scale-value depends, *inter alia*, on the mean value of Z.

The scale is limited, and the scale-value is constant only for small deviations from the normal. Consequently, in heavily disturbed areas, part of the disturbance is compensated by applying artificial magnetic fields, produced by magnets (called deflectors) put at measured distances below the needle of the field balance.

Haalck [32] has tried to unite the principles of the instruments for H and Z in his *universal balance*. It contains two magnets forming a vertical orthogonal cross, fixed to the same knife-edge. Schmidt has

PLATE 13

(a)   (b)   (c)
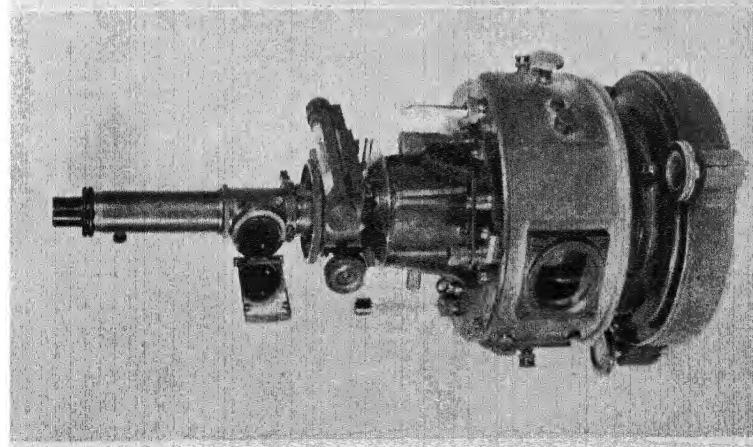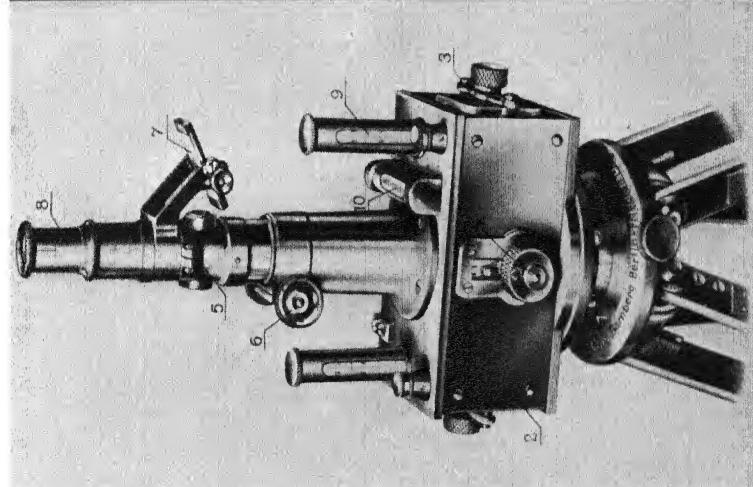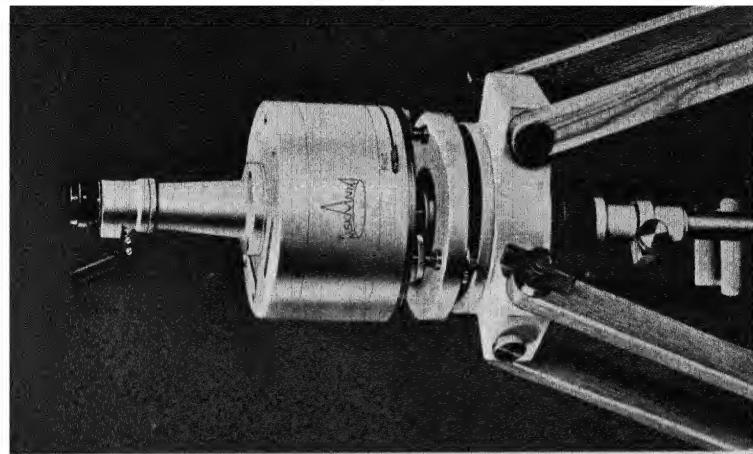
A. Schmidt's magnetic field balance : left, the original form ; centre and right, later and present forms (Hermann and Askania)

PLATE 14

c. The magnetic field balance with an attachment for temporary photographic recording (Askania)
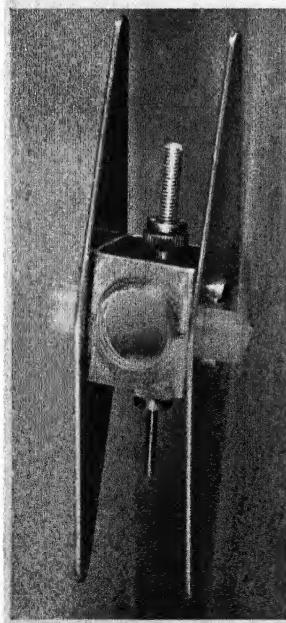
a. The magnet system (needle) of Schmidt's field balance, showing the screw adjustments (Askania)
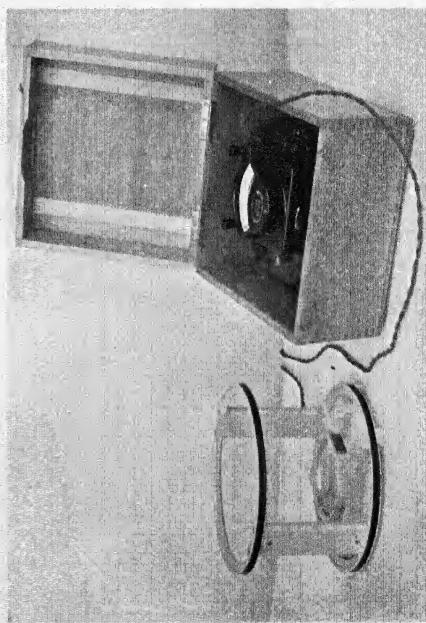
b. The coil system and galvanometer for determining the scale value of the field balance (Askania)
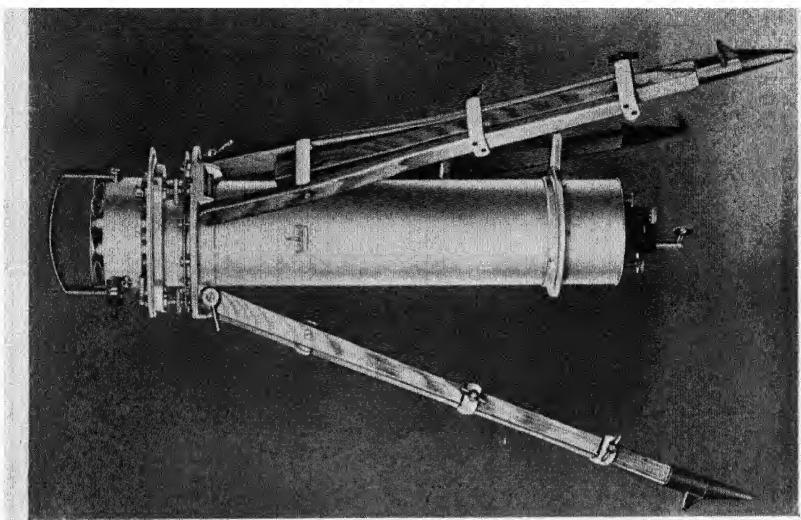
remarked that for this purpose it is better to replace the two magnets at right angles by a single magnet with slanting axis.

Another form of variometer used is the *compass variometer*, a kind of double compass (2.7), for local surveys of H.

These instruments supersede, for most purposes, the mining compasses and dip-needles formerly used, and also the Thalén-Tiberg magnetometer (1879), which measured Z by the loaded dip-needle (see 2.7) and H by deflexion of a compass-needle. The loaded dip-needle has been revived, in a new form, in the *Hotchkiss super-dip* for measuring the total intensity [18].

4.4. Technique of measurements with local variometers. Actual field-work with local variometers is simple and requires but little special training. Several precautions, however, necessary in all magnetic measurements, must be observed.

Above all, the instrument must be kept at a sufficient distance from artificial disturbances, such as iron rails and pipes, or stray electric (direct) currents. The position of any adjacent iron is important; a vertical steel post will disturb H less than Z. Roughly, an upper limit for the influence can be derived by assuming any neighbouring cast iron to be homogeneously magnetized by the earth's field (of magnitude about $0.5\ \Gamma$), giving a moment of about $10\ \Gamma$ cm.^{3} per gram. Then a disturbance of about 1γ is produced by 1 gm. of iron at a distance of 1 metre, or by 1 kg. (about 2 lb.) at 10 metres, or by 1 ton at 100 metres, and so on.

It is advisable not to measure in the immediate vicinity of the ore, where small inhomogeneities, because of their small distance from the instrument, may veil the effect of the larger but more distant masses. Measurements on light, transportable, wooden platforms of, say, 10 feet height have sometimes been made in addition to those on the surface, because the survey in the higher plane is less affected by the nearest, uppermost layer. Magnetic measurements in the mine itself are often spoiled by iron material built in, and also by the closeness of the ore.

The temperature coefficient of the instrument depends on the combined effect of changes of lengths and magnetic moments; therefore great differences of temperature before and during the measurements must be avoided in order that the thermal state may correspond to the reading shown by the thermometer, without after-effects. Even if the influence of temperature on the readings is partly or wholly compensated by mechanical or magnetic devices, the temperature of the different parts of the instrument must be equalized to about $0.1°$ C.

It has already been mentioned in 2.36 that, in northern latitudes, the needle of the Z-balance is kept horizontal by attaching the knife-edge excentrically, towards the north pole of the magnet, by less than 0·1 cm. from the mass-centre. Since the tilt is proportional to this distance, a change by only 1/1000 of its amount can affect the readings by as much as would a real change of 40γ in Z, if the mean value of Z is $40,000\gamma$. But such extremely slight displacements of mass may easily occur, perhaps by the loosening of a screw in the magnet system in the course of the observations. In order to reveal the corresponding sudden or gradual changes of the 'base-value', the plan of the survey must provide for repeat measurements at base stations at suitable intervals. If there is a difference between consecutive values measured at the same base station, and if it is not clear how to distribute the change over the stations visited in the interval, it is safer to reject that part of the series.

An accurary of between 2 and 4γ in the readings is aimed at [13 a]. The excellence of an instrument is not determined mainly by its sensitivity—which can easily be raised by mechanical or other means—but by the steadiness of its corrections for temperature, and the constancy of its base-value.

Magnetic gradiometers (2.39) have been proposed, but have as yet not been much used.

4.5. The reduction and representation of the observations. If the local anomaly is very distinct, it is, for practical purposes, sufficient to enter the readings, corrected for temperature and base-changes, upon a chart or—if the stations are arranged along a line—by a graph (called a *magnetic profile*). But if the anomaly does not exceed 200γ, say, or if magnetic storms have occurred in the course of the survey, it is necessary to eliminate the temporary variations, as observed at the nearest magnetic observatory or at a temporary observatory, as in ordinary surveys (see 2.6 and 8).

The form of the regular daily variation on normal days may be taken from Fig. 7.2. During great magnetic storms the movements of the needle are sometimes so irregular and rapid that it is safer to reject measurements made at such times; but though in sunspot-maximum years magnetic storms are more frequent than at other times, on the average there are less than two or three days per year when, at Greenwich, $\Delta D > 1°$ or ΔH (or ΔZ) $> 300\gamma$. On the other days the records of a magnetic observatory in a moderate or low latitude represent fairly well the temporary changes within about 500 km. distance; Chree found

that, on quiet days, a single observatory is sufficient for the whole of Great Britain. In regions remote from observatories, auxiliary time variometers must be put up in the neighbourhood.

The data may be plotted on a chart by entering the deviations from a suitable base-value and joining the points of equal deviation by curves called *isanomals*. If the area exceeds a few square kilometres, the normal geographical variation of the field should also be subtracted from the deviations; it may be taken from the small-scale charts of first-order surveys. In small areas of, say, 100 km. extent, it is sufficient to treat it as a linear function of latitude and longitude. In England and Germany, for instance, Z increases northwards by about 5γ/km., and H decreases by about 4γ/km.

It is not possible to eliminate the influence of topography, that is, to reduce the observations to a common level. Indeed the normal magnetic field changes very slightly with height (see 18.10), but the local anomalies change the more, the nearer the disturbing mass lies to the earth's surface. If the magnetic masses are covered with non-magnetic layers, the hilly character of the latter is not so troublesome; profiles are constructed in which the components of the disturbing forces are entered as arrows at the right height.

4.6. The determination of the magnetic properties of minerals and rocks. The interpretation of local surveys is based primarily on a knowledge of the magnetic properties (see 1.17) of the geological layers likely, or known, to be present. Fortunately, for natural minerals the magnetization I, i.e. the magnetic moment per cm.^{3}, varies between wide limits. Since it is the abnormal distribution of I inside the earth that produces the local magnetic disturbance, the survey indicates the value of I rather directly.

Samples of minerals are investigated in the laboratory by much the same magnetometric or other methods as are used in ordinary physical laboratories. If the material cannot be cut in suitable cylindrical forms, it is sometimes pulverized. The degree of accuracy is limited by the natural variability found in rock samples, and the results obtained represent the natural state only approximately, especially with regard to permanent magnetism.

It is well known that the magnetic behaviour of ferromagnetic bodies can be considerably altered by comparatively slight additions of other substances, or by thermal treatment. Thus it is uncertain how accurately such values of the magnetic constants of ferromagnetic substances as are quoted below will apply to any given specimen not

actually made the subject of magnetic measurement. In natural minerals this uncertainty is, of course, still more pronounced.

Crystalline asymmetry will not affect the field of large masses of ore under natural conditions if the single crystals are located quite irregularly. In certain exfoliated natural minerals Koenigsberger found that the susceptibility transverse to the plane of exfoliation is only 50 to 80 per cent. of that parallel to this plane; but since the susceptibility κ in these gneisses is only about 10^{-4}, the whole effect on the disturbed field will be small.

Johnson and Steiner [2.103 b] have built a rugged type of astatic magnetometer (Fig. 1) for use in the shop and laboratory to measure the susceptibility of materials. The sensitivity may be adjusted from 2×10^{-8} to 1×10^{-5} electromagnetic units per scale division, so that the volume susceptibility of practically any material may be measured. Two small cylindrical magnets ($2 \cdot 5$ mm. in diameter, 8 mm. long) are fastened rigidly together 44 mm. apart, the south pole of one below the north pole of the other, and the system, provided with a mirror reflecting a scale, is suspended by a 10-cm. quartz fibre of $0 \cdot 008$ mm. diameter. A cylindrical copper damper around the lower magnet provides nearly critical damping (that is, the needle comes to rest without oscillations). The system is made as astatic as possible; this is done as follows. The magnet system is replaced by a non-magnetic torsion-weight, and this is set at right angles to the direction of declination. Then, after again suspending the magnet system, the stronger magnet is demagnetized until the system just takes up this position again.

To calibrate a specimen or part of any geometrical shape, a hollow full-scale model of the specimen or part is made and filled with a substance of known susceptibility—for example, cupric oxide (with susceptibility $11 \cdot 0 \times 10^{-6}$). This model is placed in a fixed position as regards the magnet system and the scale deflexion is noted. The model is then removed and replaced by the specimen, and the scale deflexion is again noted; the ratio of the deflexions gives that of the susceptibilities. The sensitivity may be changed by varying the distance between specimen and magnet system. Null methods may also be adopted with this magnetometer, the deflexion being restored to zero by sending a measured current through a small Helmholtz coil around the lower magnet. The relative moments of magnets and of the magnetization of geophysical specimens may also be measured; the possible demagnetizing effect of the magnet system must then be taken into
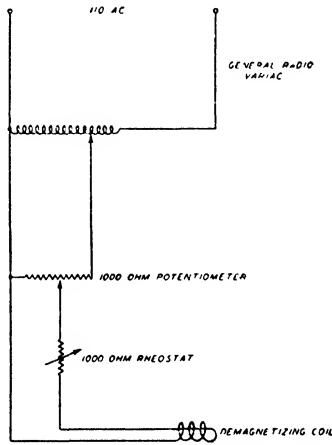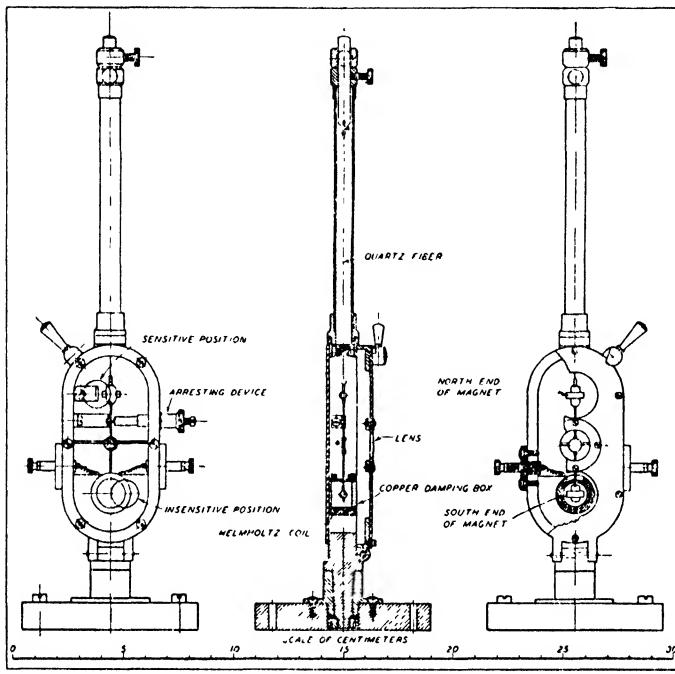
FIG. 1. An astatic magnetometer for testing the magnetic properties of materials (after Johnson and Steiner). Three views, and sketch of the circuit for demagnetizing the stronger of the two magnets in order to reach astaticism

account. By such methods it has been shown that a bar of brass may have one end diamagnetic and the other strongly paramagnetic.

Electromagnetic methods for testing the magnetic behaviour of rock

samples have also been developed by Johnson and McNish [49], with special attention to the necessity to submit the samples only to weak magnetic intensities of the order of the earth's field.

4.7. The magnetic properties of some materials. The most important magnetic mineral, and that in which magnetism was first discovered in antiquity, is *magnetite* (Fe_3O_4), for which the susceptibility κ lies between 0·1 and 30, the remanent magnetization R_0 between 7 and 50 Γ, or between 2 and 15 per cent. of I_{max}, and the coercive force H_0 between 10 and 100 Γ. Variations in the proportion of FeO to Fe_2O_3, perhaps also a slight admixture of TiO_2, may, according to Reich [8], account for the range of these values. Fine-grained magnetites have been found to show relatively higher remanence than others.

Observations show that magnetite can retain its remanent moment for long periods, though it is not certain whether it can do so over geological intervals of time. Hitherto no local magnetic anomaly has been found to change appreciably between successive surveys. Even at the observatory of Ekaterinburg, built on magnetic serpentine, where the differences between the observation pillars, only a few metres apart, reach 18′ in D, 318 γ in H, and 21′ in the dip I, an exact investigation of the series 1892 to 1912 reveals no change in the local differences.

Other strongly magnetic minerals and rocks are

    (i) Magnetic pyrites (FeS_2) .   .   . $\kappa = 0·007$ to $0·4$.

The rest of the ferriferous minerals are much less magnetic.

    (ii) Spathic iron ($FeCO_3$)   .   . $\kappa = 10^{-4}$ to 4×10^{-4}.
    (iii) Brown clay iron ore
         (limonite, $Fe_2O_3 + 2Fe(OH)_3$) . $\kappa = 10^{-4}$ to 2×10^{-3}.
    (iv) Red iron ore (Fe_2O_3)   . . $\kappa = 4 \times 10^{-5}$ to 3×10^{-3}.

Eruptive minerals are all more or less magnetic:

    (v) Serpentine ($Mg_3Si_2O \cdot 2H_2O$) . $\kappa = 0·0006$ to $0·014$
        (this has a remanent magnetization 0·6 Γ).
    (vi) Basalt .   .   .   .   . $\kappa = 0·0002$ to $0·01$
                                   (Northumberland).
    (vii) Mean of 45 dolorites and basalts from the British Isles, according to Rücker   .   .   . $\kappa = 0·0026$.
    (viii) Granite   .   .   .   . $\kappa = 10^{-5}$ to 10^{-3}.
    (ix) Gabbro   .   .   .   . κ up to $0·005$.
    (x) Lava from Mount Etna .   .   . $\kappa = 0·008$.

Basaltic rock pillars, if struck by lightning, acquire a very strong

permanent magnetization, the polarity of which indicates the direction of the electric discharge.

Sediments have generally very low values for κ. Quartz, chalk, dolomite, clay, and coal have susceptibilities κ between 10^{-6} and 10^{-5}. Rock salt is diamagnetic (κ about -10^{-6}), and so also is graphite (κ about -10^{-5}).

Though the maximum susceptibility κ_{max} of natural rocks is lower than that of pure minerals, their coercive force H_0 is relatively high. To a certain extent the ratio H_0/κ_{max} measures the resistance offered to demagnetization. It is therefore important to know that in the eruptive rocks investigated by Turcev [8] this ratio is of the order 100 to 1,000, much higher than for tempered steel (6·0) or pure magnetite crystals (0·5).

Eve and Keys [4] give two examples illustrating the wide range of magnetic behaviour in natural substances. The magnetite at Caribou, Colorado, contains titanium; its coercive force is so strong that each splinter of the ore, when hung by a thread, behaves like a compass-needle. The magnetite at Mineville, N.Y., resembles soft iron, changing its polarity when simply rotated from a horizontal to a vertical position in the earth's magnetic field.

With respect to the origin of magnetic anomalies, it is perhaps of importance that certain minerals increase their susceptibility up to twenty times the initial value after they are heated as high as 700°. The analysis of transparent sections shows that this change in susceptibility is accompanied by changes in the mechanical texture. This bears on the strong magnetization observed in lava flows. For a more detailed account of the magnetic behaviour of natural rocks see, for instance, the report by Koenigsberger [3.22 b].

4.8. The general connexion between geological structure and magnetic anomalies. Nippoldt [12] has drawn a chart of Europe showing the magnetic anomalies by the isanomals of the vertical intensity. The smooth run of the isomagnetic lines is completely marred in a large region adjacent to and comprising the Baltic Sea, i.e. in Finland, Scandinavia, Denmark, East and West Prussia, Northern Poland. Other disturbed regions are Roumania and Bessarabia, parts of Belgium, the Auvergne, Scotland, and the Hebrides. The Alps and the Transylvanian Alps show less intense disturbances of varying sign.

The discussion of this chart, together with the geological considerations outlined by Reich, give a rather clear picture of the geological

cause of local anomalies, consistent with the laboratory experiments on minerals.

The crystalline rocks, which in Finland and Scandinavia come to the surface, cause strong magnetic anomalies. Their influence reaches to the surface even where they are covered by non-magnetic layers—as, for example, by about 100 metres of water in the Baltic Sea, or by mesozoic and younger layers of more than 1,000 metres thickness in eastern Prussia.

If these anomalies were produced merely by *induction* by the present geomagnetic field in highly susceptible rocks, southern magnetism would be induced at the surface, and Z at the surface would be above the normal. The existence of some strong negative anomalies therefore supports the conclusion that magnetic rocks may be *permanently* magnetized in directions differing from that of the present field (see also § 12). It is uncertain whether this permanent magnetization was produced by the general field of the earth at the time of congelation or metamorphosis, or whether other causes must be considered; all permanent magnets are observed to lose their magnetism gradually, but experience does not suffice to indicate the loss of moment in geological times.

Younger eruptive rocks also cause intense anomalies, but except in the Auvergne (north-magnetic) and the Hebrides (south-magnetic) they cover rather small areas. On the basaltic Vogelsberg in Hesse, Z exceeds the normal value by more than $0{\cdot}01\ \Gamma$ over an area of 300 km.[2]

The tertiary mountain chains of Europe show disturbances of which the intensity varies according to the degree in which crystalline rocks participate in the geological structure.

4.9. The technique of magnetic prospecting. In all cases where differences of susceptibility or permanent magnetization between some useful mineral and the surrounding rocks are to be expected, magnetic methods of prospecting may be applied. The first aim of the survey is to find the distribution of magnetization inside the earth. Unfortunately this problem is completely indeterminate, and has an infinite number of solutions, even if the magnetic field were known everywhere outside the magnetic masses, and not only, as in the survey, at a surface; consequently the interpretation of magnetic surveys is to some extent guess-work. Usually the geological structure, so far as it is known, will restrict the possibilities. For simple magnetic distributions, such as cylinders, plane layers of different inclinations, dislocations, etc., the magnetic field has been calculated, assuming homogeneous magnetiza-

tion; the results have been summarized by Haalck [7] in formulae and characteristic profiles. Owing to the rather rough nature of prospecting work, the effect of demagnetization (p. 24) has not been taken into account, so that the diagrams are only schematic. The comparison of
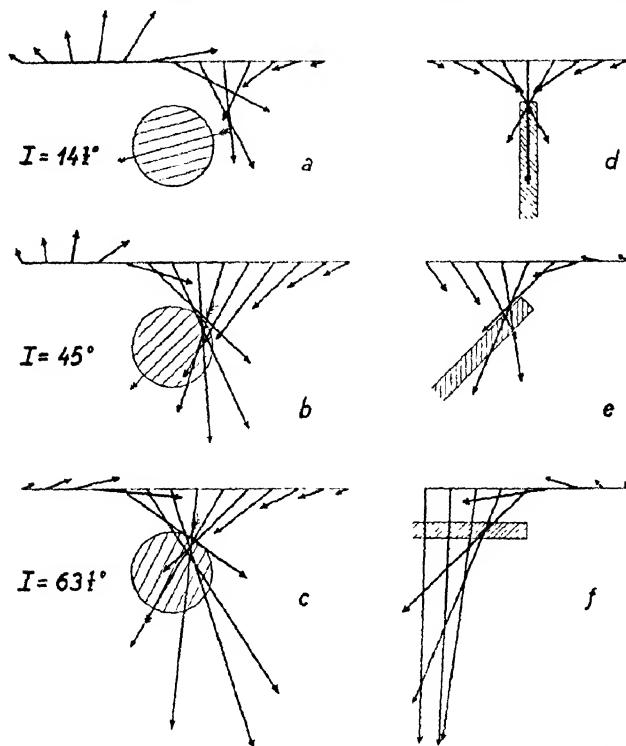
FIG. 2. Profiles along a magnetic meridian showing the vectors of the magnetic anomaly produced by a sphere (left) of high permeability for three values of the magnetic inclination I, and also showing profiles perpendicular to the magnetic meridian for a thin layer (right). (After Haalck; a more complete set of examples has been given by Gulatee [7 a])

these ideal cases with the results of the survey will facilitate the interpretation. Haalck's results have been corrected and supplemented by Gulatee [7 a].

Many anomalies investigated hitherto have justified the working hypothesis that the additional magnetic moment has been acquired by induction by the present normal magnetic field of the earth.

In Fig. 2, a–c, the disturbing effect of a homogeneously magnetized sphere is shown for three values of the magnetic inclination I; the profiles are drawn in the plane of the magnetic meridian. Numerical values can be taken from the formulae given in 1.15. The profiles normal to the meridian are symmetrical, and similar to those for $I = 90°$, with

the biggest disturbance exactly over the centre of the sphere. For $I > 65°$, as in England, the profiles do not depend very much on their magnetic azimuth.

Fig. 2, *d–f*, shows the effect of a thin plane layer, inclined at different angles, and stretching to infinity in one direction. The upper border of the layer is supposed to be in the magnetic meridian, and profiles are drawn perpendicular to it.

A fault in a horizontal magnetic layer of finite thickness is the more
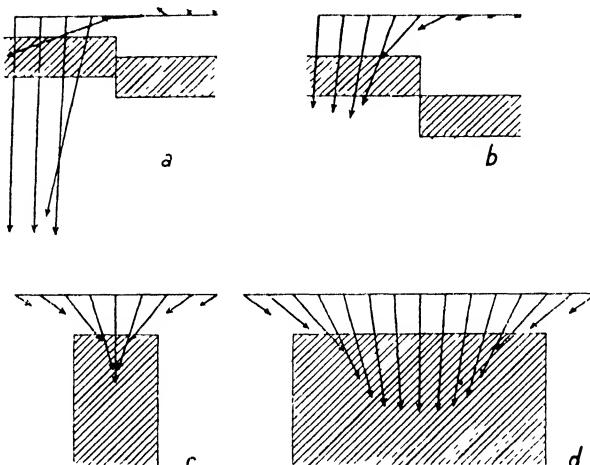
Fig. 3. Profiles showing the magnetic anomalies across faults and vertical layers of finite thickness. (After Haalck)

conspicuous magnetically the nearer the layer is to the surface (Fig. 3, *a, b*). Finally, the influence of a plane vertical layer of finite thickness is considered (Fig. 3, *c, d*) for the values 2 and 6 of the ratio: thickness/(depth of upper border). The increase of the central disturbance with growing thickness is not very marked.

Of all the magnetic elements, the vertical intensity Z is most simply connected with the position of the disturbing mass (in regions where the inclination is greater than 45°). The horizontal component of the disturbance appears partly in D, especially at the eastern and western limit of the mass, and partly in H, especially at the northern and southern limit. The diagrams show, in general, a maximum or minimum of the horizontal component of the disturbing force in regions where the vertical component varies most rapidly, and vice versa.

The intensity of the disturbance at the surface decreases with the depth d of the disturbing mass: for instance, as $1/d$ in the case of a vertical unlimited layer, or as $1/d^3$ in the case of a sphere.

Rössiger and Puzicha, in an important paper [23] on the mineralogical, geological, and geomagnetic aspects of anomalies in the Harz, have pointed out the appreciable influence of demagnetization in natural anomalies over *highly* magnetized ores. They find that the magnetic field of plane strata is most easily calculated if the magnetization is considered as the result of two components, the one parallel, the other perpendicular to the plane; their formulae are most convenient for the interpretation of local magnetic surveys. Two simple rules may be cited:

(*a*) If a magnetic anomaly ΔZ (expressed in γ) is caused by a steep extended layer of rock containing magnetite with the concentration v per cent. by volume, then approximately $v = \Delta Z/700$, so that a concentration of magnetite of no more than 1 per cent. by volume produces an anomaly of 700γ.

(*b*) The susceptibility κ of such a layer can be estimated from the anomaly ΔZ and the average value of Z by the relation $\Delta Z = 2\pi\kappa Z$. An anomaly $\Delta Z = 1,000\gamma$ in a region where $Z = 42,000\gamma$ indicates that $\kappa = 0 \cdot 0038$.

Even if the magnetic survey does not yield definite conclusions about the geological structure, it may serve to indicate points where borings should be made, while, in their turn, the borings will help in the right interpretation of the survey.

4.10. Methods for calculating the field of magnetized bodies.

There are only a few simple geometric forms of disturbing bodies for which the magnetic field can be described by not too cumbersome formulae. In order to evaluate the field of any configuration of different layers, supposed to be homogeneously magnetized, Haalck [7] has developed a graphical method, of which the leading idea (taken from similar methods in torsion-balance surveys) may be illustrated in a simplified case.

Imagine a geological structure which does not change in the magnetic east–west direction, so that it is characterized by a single geological profile running magnetically north and south. This profile may show a number of layers, all homogeneously magnetized in the plane of the magnetic meridian, but with different intensity and direction. Consider, for this two-dimensional case, a certain point A above the surface. That part of the magnetic field at A, which originates in a single layer, is the sum of the magnetic actions of all particles, or of each square centimetre, of the area covered by the layer in the profile.

Generally, this sum is actually formed by ordinary integration for

the vertical and horizontal component of magnetization separately. The graphical method, however, consists in dividing the whole area, relatively to A, into a great number of small rectangles, the size of which is adjusted so that each contributes the *same* amount to a certain field component in A. Diagrams, dividing up the whole plane into such elementary rectangles, can be drawn once for all; naturally the size of the rectangles increases with the distance from A. The geological profile is drawn on tracing-paper, placed over the diagram (graticule), and the number of rectangles in each particular layer is counted and multiplied by the appropriate component of magnetization. By shifting the two drawings, the field can easily be got for as many points A as desired.

The foregoing explanation gives only a rough idea. The complete evaluation requires two such diagrams, and the theory is based on the relations between the gravitational and the magnetic field of disturbing homogeneous bodies (see § 13).

Nippoldt [9] has worked out the field of various aggregates of magnetic poles for use in prospecting, especially for determining the depth of a disturbing body or layer. Jenny [34] imitates the anomaly on a small-scale model by magnets.

4.11. Examples of local magnetic anomalies. The investigation of the underground position of deposits by magnetic methods is of course most successful in the case of *magnetite*.

The biggest anomaly in the world is that of Kursk, about 400 km. south of Moscow. It is confined to two narrow strips, running in parallel, from NE. to SW., at about 60 km. distance apart. The northern strip is the more intensely disturbed; it is 250 km. long, but, where the disturbance is most extreme, only 2 km. broad. The vertical intensity is everywhere above the normal value, and ranges up to $1·9\ \Gamma$, five times the normal value there. The horizontal disturbance vectors are directed towards the axial line (Figs. 4–6). Borings revealed magnetite-quartzite containing 40 per cent. of iron at about 150 metres depth; the thickness of the layers varies up to 400 metres, and they are inclined at a steep angle. This large mass of iron is, however, of restricted commercial value because the admixture of quartzite prohibits the working of the ore.

Haalck [27] has applied his graphical method (see § 10) to the Kursk anomaly. He concludes that the layers found by boring might explain the anomaly in so far as their magnetization can be taken as parallel to the normal magnetic field, but the specific moment (about 0·7 gauss) must be three to four times higher than the normal field could produce

by induction. Rejecting (perhaps wrongly) the hypothesis that the ore could retain a permanent magnetic moment throughout geological intervals of time, he infers that the real cause of the Kursk magnetic and gravimetric anomaly is a deposit of magnetite at a depth of 500 to 1,000 metres. This ore would have a strong magnetic moment, induced
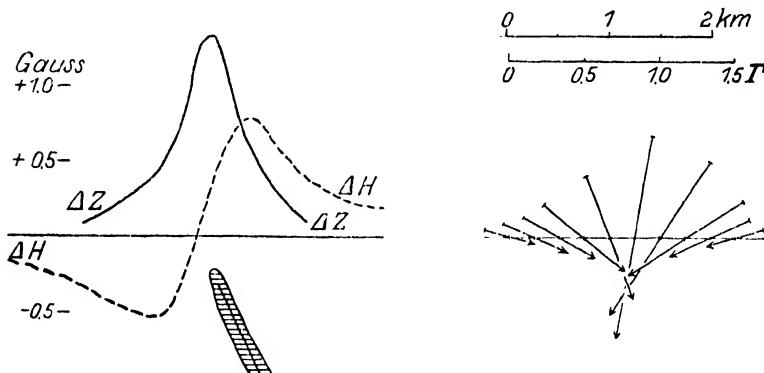
FIG. 4. Typical profile across the disturbed strip of the Kursk anomaly; curves (left) and vectors (right)
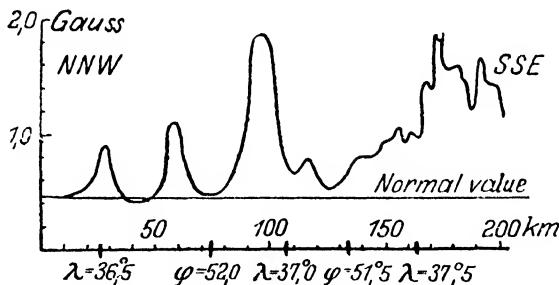
FIG. 5. Profile for the vertical intensity Z along the axial line of the Kursk anomaly (points where the profile cuts circles of latitude ϕ, or meridians of longitude λ, are indicated below)

by the present normal field, and the strong field of these deeper layers could induce the magnetization of the higher quartzite layers.

The big disturbance of Kiirunavaara in Lapland (Fig. 7), investigated by V. Carlheim-Gyllensköld, has a similar elongated form. Since the ore extends to the surface, its action is stronger than at Kursk; Z rises to $3 \cdot 6 \ \Gamma$. The thickness of the deposit varies between 28 and 164 metres; it contains almost pure magnetite. The magnetization, as at Kursk, is parallel to the present direction of the normal field vector, and its intensity (about $0 \cdot 3 \ \Gamma$, with a susceptibility of $0 \cdot 8$) is not higher than the present normal field could induce [60].
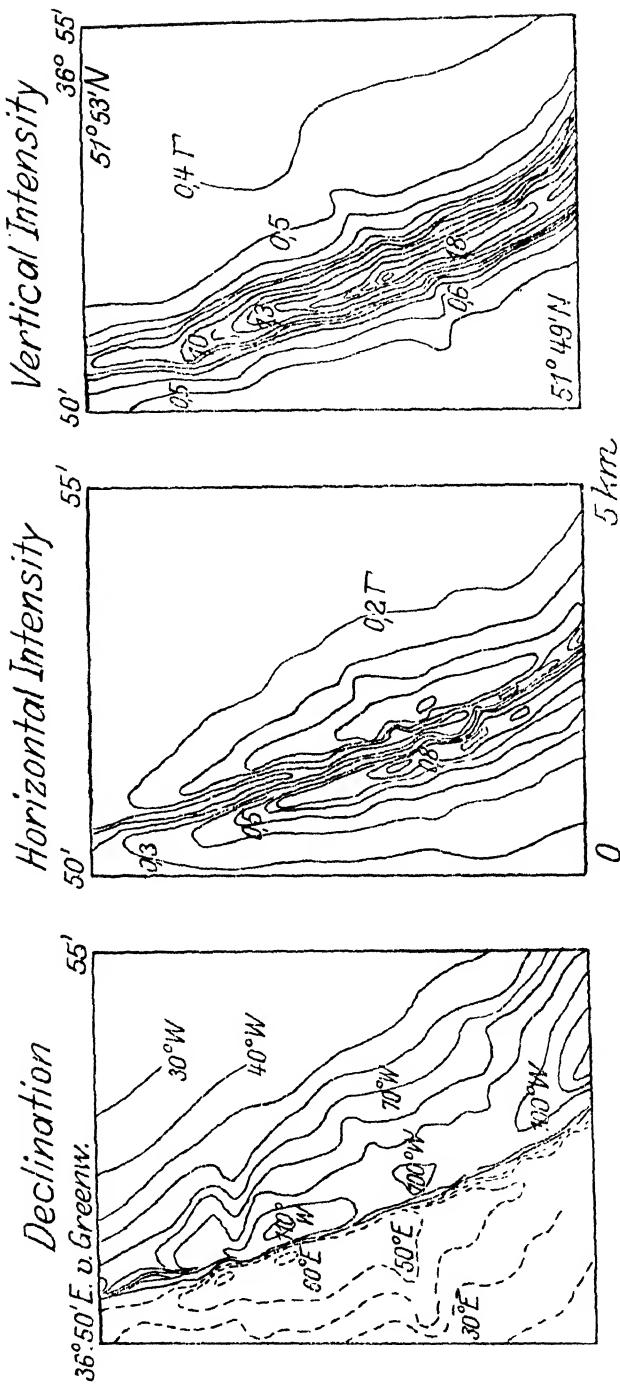
FIG. 6. Isomagnetic charts of the northern strip of the Kursk anomaly. (After Lasareff)

A similar anomaly caused by iron ores is that of Krivoy Rog in southern Russia. Near the Finnish island of Jussarö in the Baltic Schären, a deposit of magnetite increases Z by 63,000 γ, and the disturbance in declination has caused many shipwrecks. In Germany the relatively small deposit at Berggiesshübel in Saxony was surveyed by
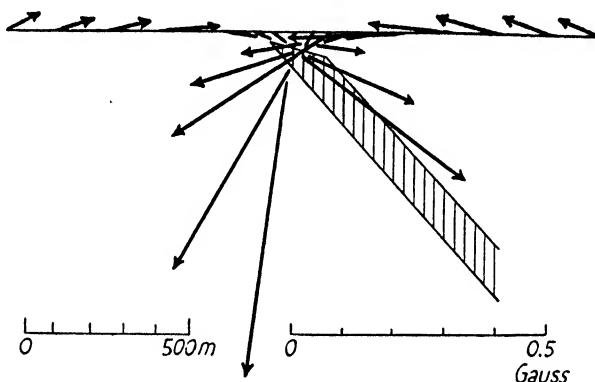
FIG. 7. Disturbance vectors for a profile across the iron-ore anomaly at Kiirunavaara, Lapland. (After Carlheim-Gyllensköld)

Heiland [11]; the observations of H and Z enabled him successfully to indicate a spot where an adit disclosed a magnetite dyke (Figs. 8, 9).

The sedimentary deposits of iron ore in Northamptonshire, investigated by Cox, produce only slight disturbances of about 100 γ.

Heiland [5] has discussed the results of magnetic prospecting work in North America. Frequently, not the ore itself, but the formation containing or accompanying it, is located. In the Lake Superior district, where a heavy burden of glacial drift veils the substructure, both iron and copper are traced by magnetic measurements. Concentrations of gold in river channels have been successfully located by magnetometric methods, because the noble minerals are usually associated with heavy magnetite.

Magnetic prospecting for oil and for salt domes requires sensitive instruments and careful observations. In Germany negative anomalies in Z of the order 50 γ were found over salt domes. The slight diamagnetism of the salt compared with the para- or ferromagnetism of its surroundings does not sufficiently explain these anomalies; their cause may lie at greater depths in the bed-rock. On the Gulf Coast of Texas magnetic prospecting for salt domes has, however, been abandoned in favour of seismic or gravimetric methods.

The economic value of oil, compared with which the costs of **geo-**

physical surveys are negligible, has led to great prospecting activity in North America. According to Heiland [5], the chances for *magnetic* prospecting are most favourable in cases where the oil is accumulated along the fissures, or intersections of fissures, accompanying volcanic dikes, because these are indicated by relatively high values of Z.
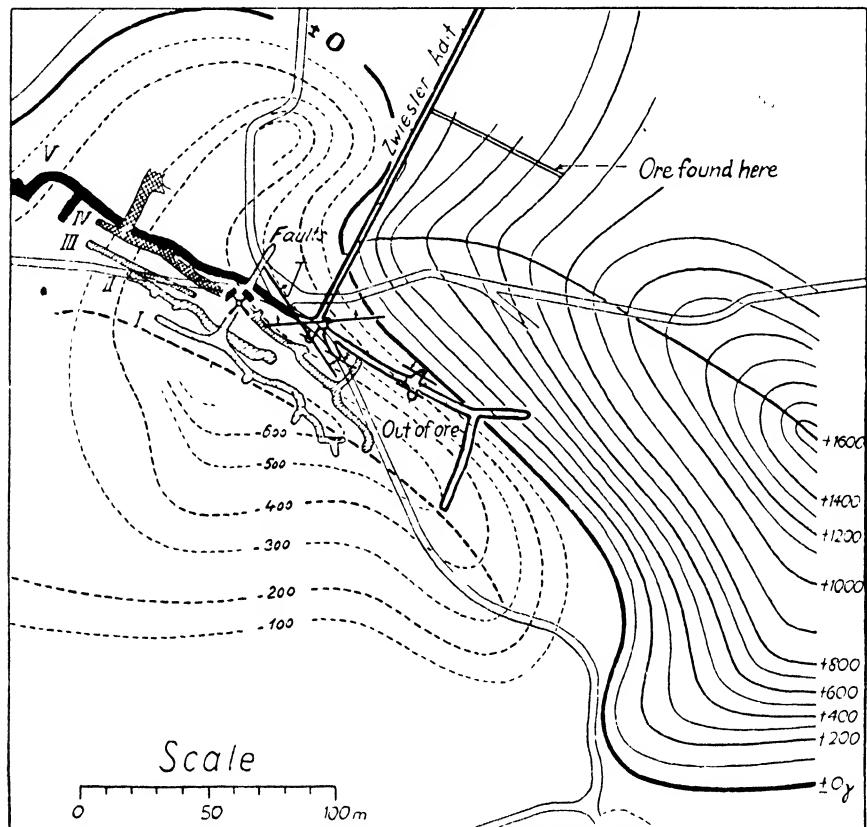
FIG. 8. Isanomals of the vertical intensity above the magnetic ore deposit of Berg-giesshübel, Saxony. (After Heiland)

4.12. Induced or remanent rock magnetism?

It has already been mentioned that many strongly magnetic ores are magnetized in a direction quite compatible with the present geomagnetic field. This justifies, to a certain extent, a tendency to attribute all local disturbances to induction by the field. However, it is certain that igneous rocks acquire a remanent magnetization during their cooling when the temperature passes through the Curie point; the magnetic elementary particles, free to move in the molten lava, aline themselves in the direction of the magnetizing field, and keep that direction when the
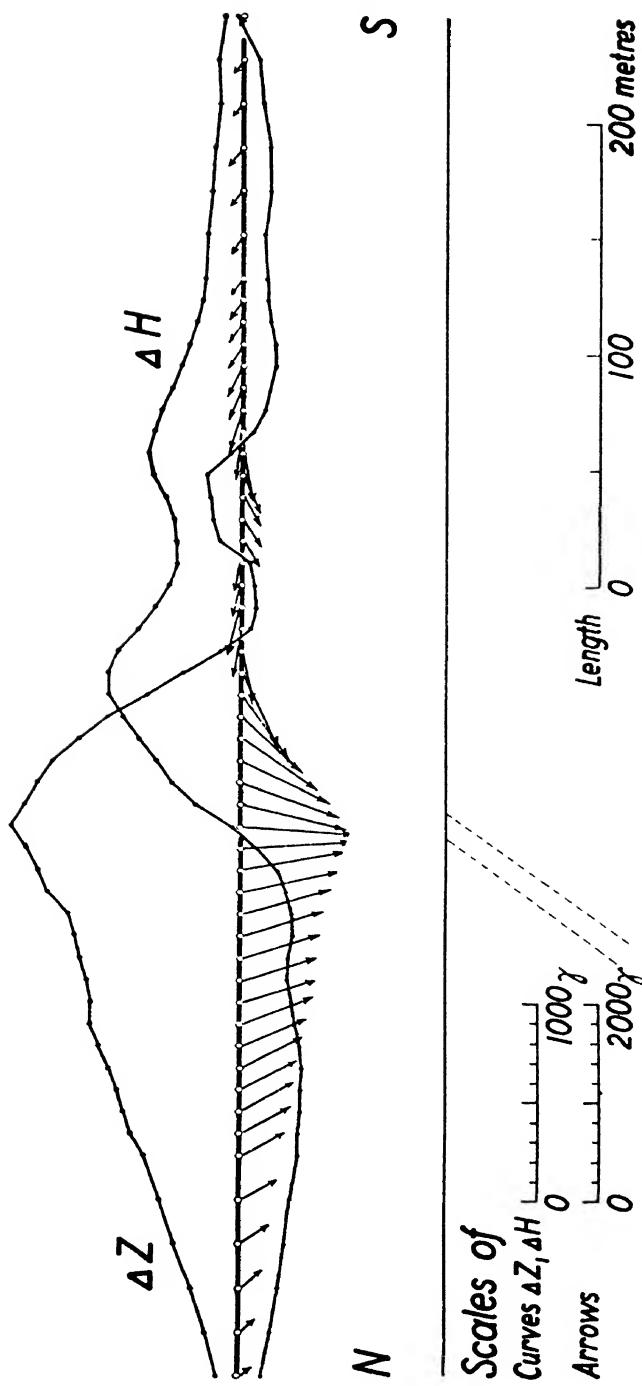
Fig. 9. North–south profile across the Berggiesshubel deposit, showing (below) the approximate form of the ore deposit and (above) the horizontal and vertical anomalies, and the total disturbance vectors for the anomaly

rock solidifies. Opinions are divided as to whether this remanent mag-
netization could be retained through the millions of years that have
elapsed since the formation of the igneous rocks, even if the magnetic
field changed its direction considerably. Local anomalies of abnormal
polarization, that is, having a magnetization opposed to the present
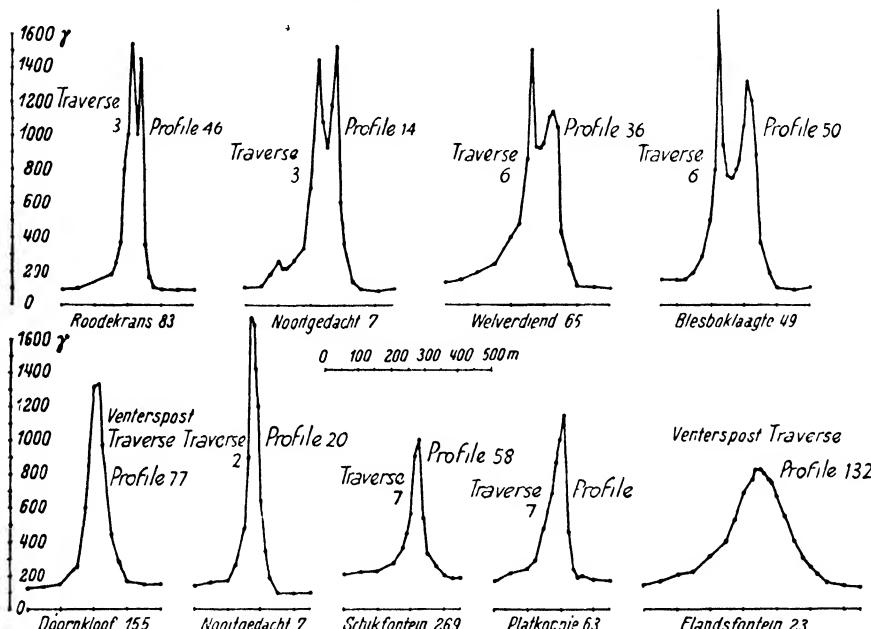
FIG. 10. Profiles of the vertical intensity across the Pilansberg system, South Africa.
(After Gelletich)

field, have been cited. Heiland [50] has discussed possible causes for
such 'negative anomalies', compatible with the view that the remanent
magnetization was acquired by induction by a general magnetic field
having the same direction as to-day: for instance, opposite magnetiza-
tion induced in smaller bodies by larger bodies in the vicinity, the over-
turning of permanently magnetized formations as a whole, or other
mechanical effects.

There is, however, an example of a widespread palaeozoic system of
volcanic dykes in South Africa, extending over an area 300 km. long
and up to 150 km. wide, which seems to defy any attempt at an ex-
planation other than that of a complete reversal of the magnetic field
in that region at the time of the formation. The Pilansberg, 120 km.
north-west of Pretoria, sends out more than fourteen dikes, of about
100 miles in length, spread in a fan-shaped manner. They are more or

less vertical and plate-like; their thickness varies between 20 and 100 metres. They are described as alkalisyenitic and doleritic, with uniform mineral composition. Gelletich [30] found that the whole system —the biggest dike system ever described—is everywhere magnetized in a direction *opposite* to the present magnetic field; they cause anomalies in Z of between 200 and 1,700γ (Figs. 10, 11). This 'fossil magneti-
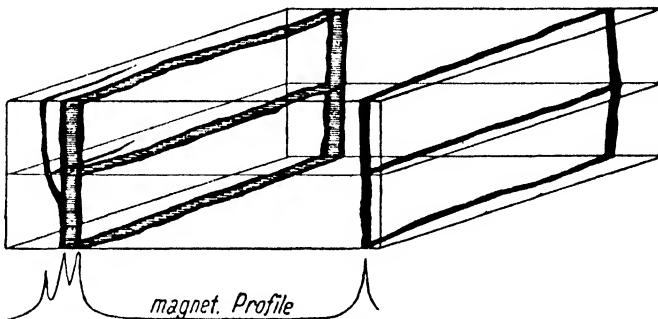
FIG. 11. A schematic view of volcanic vertical dikes, with a magnetic profile of the anomaly in the vertical intensity

zation' seems to have survived 200 million years. Geological bodies in the vicinity, rich in magnetite, formed before and after the Pilansberg system, show a magnetization conforming to the present-day field.

4.13. Relations of magnetic to other geophysical prospecting methods. As regards theory and interpretation, magnetic prospecting for bodies with induced magnetism is closely related to the two methods depending on the field of gravity, in which measurements are made either with pendulums, which give the actual force of gravity, or with the torsion balance, which gives the horizontal gradients and curvature of gravity. For bodies of homogeneous mass-density ρ and of homogeneous magnetization μ in the direction m, the gravitational potential U and the magnetic potential W are connected by the equation (G = gravitational constant = $6 \cdot 7 \times 10^{-8}$ c.g.s.)

$$W = -(\mu/G\rho)\, \partial U/\partial m.$$

Thus the magnetic force ($-\mathrm{grad}\, W$) is proportional to the gradient of that gravitational component which is parallel to the direction of magnetization. Using this relation, Eötvös derived formulae connecting the components of the magnetic field with the gradients and curvatures of gravity measured with the torsion balance; but no practical application of them seems to have been made till Haalck [27] developed his graphical method mentioned above (see also a paper by Nagata [56]).

Since magnetic minerals are generally of higher density than others, the magnetic anomalies may be expected to be connected with those of gravity. Indeed the Kursk anomaly, along the axis of the disturbance, shows values of gravity g which exceed the normal value (981·226 cm./sec.^{2}) for that region by 0·087 cm. sec.$^{-2}$ and more. Likewise in Central Europe the principal anomaly of gravity, parallel to the tertiary mountain ranges, is disturbed magnetically also (Nippoldt [12]); but as far as pendulum observations are available, several magnetic anomalies, for instance in eastern Prussia, are not accompanied by disturbances of gravity.

This practical looseness of the connexion is easily explained: the ratio of the densities of the heaviest and lightest minerals is less than 10, whereas the ratio of the magnetization may attain 10^7 and more. Moreover, the rate of decrease, with increasing distance, in the gravitational attractive *force* of a mass, is less, by about one power of the distance, than for the *gradients* and *curvature* of the field of gravity, or for the *magnetic* force; consequently pendulum observations show a relatively greater influence of the deeper layers than torsion-balance and magnetic measurements.

For seismic, electric, and other geophysical prospecting methods of applied geology, as well as for more detailed accounts of the instrumental and geological and mineralogical aspects of magnetic prospecting, the reader is referred to the books and papers cited. In order to prevent misunderstanding it may be mentioned that natural earth-currents, so far as they are caused in the earth's outer crust by the electro-chemical action of different minerals (especially in sulphide ores), are too small to produce measurable magnetic disturbances at the surface (Ch. XIII).

From the practical standpoint, magnetic prospecting has an advantage over other methods because of its relatively small cost.

V

SOLAR AND LUNAR DATA

THE earth's variation-field ΔF (apart from the secular variation) is found to depend upon the relative positions of the earth, sun, and moon, and also upon the intrinsic physical state of the sun. The geometrical relationships between the three bodies are known with great accuracy, and can be predicted; our knowledge of the sun's physical condition, and of what are the particular features which most affect the earth's magnetic field, is much less complete.

5.1. The earth's orbital motion. Over such periods of time as will be considered in this book, the directions of the axes of rotation of the earth and sun may be considered as fixed (relative to the stars). The two planes normal to them, through their centres, are the equators of the earth and sun, which are likewise fixed in direction.

In the course of each calendar year the earth describes a complete orbital circuit round the sun in a plane called the ecliptic. The linear distance between the centres O and C of the earth and sun varies by 1·5 per cent. on either side of the mean distance, which is $1·495 \times 10^{13}$ cm. or approximately 150 million km.; the maximum and minimum distances are attained a few days after July 1 (aphelion) and a few days after January 1 (perihelion) respectively; at the latter date the earth receives about 6 per cent. more heat radiation from the sun than in July. The inclination of OC to the earth's equator (or the declination δ of the sun) attains its northern and southern maxima of about $23\frac{1}{2}°$ about June 21 (the June solstice) and December 21 (the December solstice): it is zero at the equinoxes, about March 22 and September 22. The sun's equator is inclined at $7·3°$ to the ecliptic or plane of the earth's orbit: the two planes intersect along a line through which the earth passes on about June 5 and December 5, so that on these dates the earth is in the plane of the sun's equator. Between June 5 and December 5 the earth is north of the sun's equator, its heliographic latitude (that is, the inclination of CO to the sun's equator) attaining its maximum value of $7·3°$ on about September 5; on about March 5 it attains its southern maximum.

All these changes—of the distance OC, the sun's declination, and the earth's heliographic latitude—recur annually, in $365\frac{1}{4}$ mean solar days, or $366\frac{1}{4}$ sidereal days. For meteorology the seasonal changes depending on the sun's declination are the most important, but it cannot be

assumed without examination that the same applies to the earth's magnetic changes.

In considering the seasonal variations of the earth's magnetic field,
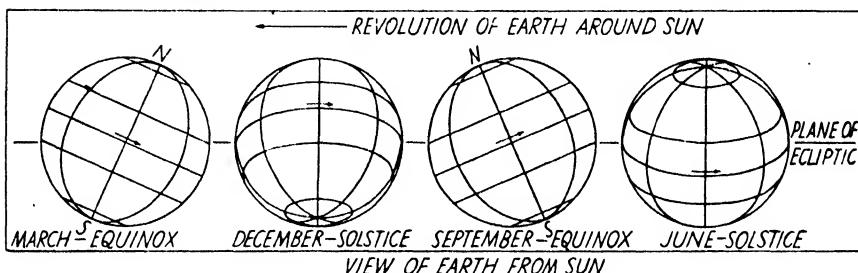
FIG. 1. Views of the earth from the sun, at different seasons
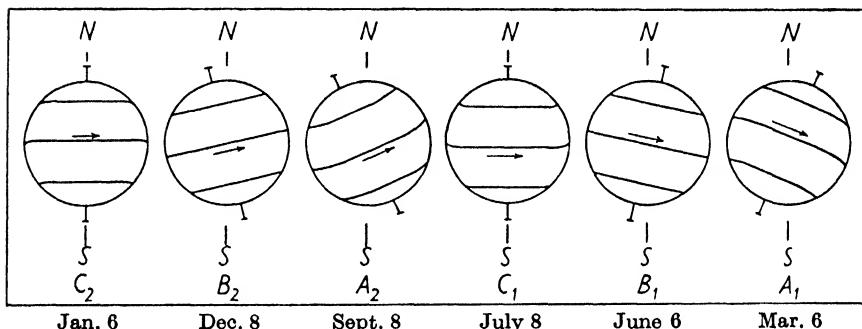
FIG. 2. Views of the sun, at apparent noon at different times of the year, from a point P on the northern hemisphere of the earth. NS shows the direction of the local (vertical) meridian plane at P. On the days A_1, A_2 the sun's axis is most tilted from the perpendicular to our line of sight, and we then see the maximum amount of the northern (or southern) hemisphere. On the days B_1, B_2 the sun's axis is perpendicular to our line of sight. On the days C_1, C_2 the sun's axis appears vertical at noon

the twelve months of the year will usually either be treated separately or grouped into three sets as follows:

June solstitial group: May, June, July, August.

Equinoctial group: March, April, September, October.

December solstitial group: November, December, January, February.

These may be referred to by the letters *j*, *e*, *d*.

A list of the main annual events relative to the positions and orientations of the earth and sun, and referring to the points and sketches in Figs. 1, 2, and 3, is given in Table 1 for the year 1934; for other years, similar data can be found in the 'Ephemeris for Physical Observations of the Sun' given in the Nautical Almanac, or in the American Ephemeris. They give the position angle P made by the sun's axis with the terrestrial meridian half-plane which passes through the sun's centre (P is
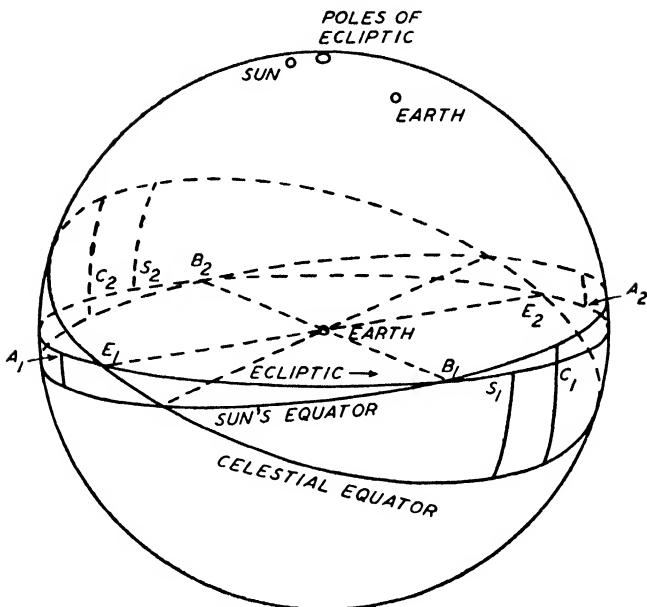
FIG. 3. Celestial globe showing the relative positions of the ecliptic, the sun's equator, and the celestial (or earth's) equator. The letters are explained in Table 1. The north poles of the ecliptic, of the sun, and of the earth, are indicated at the top

TABLE 1. *Various epochs for the year* 1934

Point	Date	
...,	Jan. 2.	Perihelion; the sun's apparent semi-diameter is 16′ 17·6″.
C_2,	Jan. 6.	The sun's axis lies in the terrestrial meridian plane through the sun's centre C ($P = 0$).
A_1,	March 6.	The greatest southerly value of the earth's heliographic latitude ($D = -7° 15′$).
E_1,	March 21.	The spring equinox (the sun's declination δ is zero).
...,	April 7.	The greatest westward angle of the sun's axis to the meridian ($P = -26° 25′$).
B_1,	June 6.	The earth's heliographic latitude is zero ($D = 0$).
S_1,	June 22.	The June solstice (maximum northern declination of the sun, δ = 23·5°).
...,	July 5.	Aphelion; the sun's apparent semi diameter is 15′ 45·4″.
C_1,	July 8.	The sun's axis lies in the meridian. ($P = 0$).
A_2,	Sept. 8.	The greatest northerly value of the earth's heliographic latitude ($D = +7° 15′$).
E_2,	Sept. 23.	The autumnal equinox (δ = 0).
...,	Oct. 11.	The greatest angle of the sun's axis to the east of the meridian ($P = +26° 25′$).
B_2,	Dec. 8.	The earth's heliographic latitude is zero.
S_2,	Dec. 22.	The December solstice (maximum southern declination of the sun, δ = -23·5°).

reckoned positive if the north pole of the sun's axis is inclined towards the eastern side of the meridian), the heliographic latitude D of the centre of the sun's disk as seen from the earth, and the heliographic longitude of the centre of the sun's disk, according to Carrington's elements (see §4).

The *Julian Day-Number* is reckoned consecutively from January 1, 4713 B.C. For instance, 1939 January 1 has the Julian Number 2429265. More precisely, this integer number refers to the mean Greenwich noon of the day; for instance, 2429265·25 refers to 6 p.m. or 18^h G.M.T. of that date. The astronomical year-books, nautical almanacs, and ephemerides give tables for the Julian Numbers, which are convenient for many purposes, e.g. for research on periodicities (Schmidt [12.20]; Pollak [16.21]).

5.2. The month and the lunar day. The moon describes an orbit round the earth in a plane inclined at $5·1°$ to the ecliptic; the pole of the orbit revolves about that of the ecliptic once in 18·6 years, so that the inclination of the plane of the moon's orbit to the earth's equator varies between $23·5° \pm 5·1°$ or $18·4°$ and $28·6°$. The moon's declination consequently changes during each passage round its orbit between maximum northern and southern values which may vary from $18\frac{1}{2}°$ (for instance, in the year 1941) to $28\frac{1}{2}°$ (in the year 1950); in midwinter, therefore, the full moon will stand $10°$ higher at midnight in 1950 than in 1941.

The mean distance of the moon from the earth is 384,400 km., or 60·27 times the earth's radius. The eccentricity of the orbit is considerable, and slightly variable; the mean ratio of the maximum distance, at apogee, to the minimum value, at perigee, is 1·1162, and the maximum ratio is 1·1411. The apogee revolves round the lunar orbit once in 8·8 years.

The moon's passage round the earth is accomplished in $27^d\ 7^h\ 43^m$ (the sidereal period), so that the mean lunar day, or average interval between two successive passages of the moon across any terrestrial meridian, is $24^h\ 50·47^m$, though the actual interval varies owing to the changing distance and orbit of the moon.

The moon revolves round the earth relative to the line OC through the sun's centre once in 29·5306 days, a period which is called a lunation, or the *synodic* or lunar month. The moon's phases depend on the angle ν between the meridian half-planes through the sun and moon (reckoned positive when the moon is to the east of the sun). When $\nu = 0$ there is new moon, and the values of ν at first eighth phase, first quarter (half-moon), full moon, and so on, are $45°$, $90°$, $180°$, and so on.

The east longitude τ' of any station P relative to the meridian oppo-site to that containing the moon is a measure of the local apparent lunar time. Clearly, if t' is the local solar time (Fig. 4),

$$t' = \tau' + \nu \qquad (1)$$

if angular measure is used.

Mean—and not apparent—solar time t is used (as in civil life) in calculating solar daily variations, because the use of mean time is
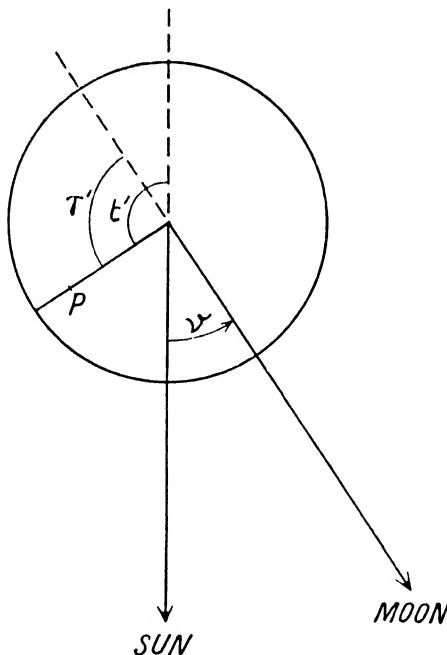
FIG. 4. The earth as seen from above the north pole

more convenient in observatory work, and because the difference be-tween apparent and mean time can be readily allowed for in the dis-cussion. By analogy with the mean solar time t, a mean lunar time τ is introduced, determined by the motion of a fictitious mean moon, which is imagined to revolve uniformly round the earth. Just as t is counted in 24 solar hours from midnight to midnight, τ is 0 for the lower transit of the mean moon and increases by 24 lunar hours up to the next lower transit. A mean lunar day equals 1·03505 solar days, or 24^{h} $50·47^{\text{m}}$. If we re-define the angle ν introduced above as the angle between the meridian half-planes containing the *mean* sun and moon,

$$t = \tau + \nu, \qquad (1\,\text{a})$$

where t, τ, and ν may be reckoned either in angular measure or in hours ($15° = 1$ hour). The *age* ν of the moon completes a full cycle of 24 hours in the course of a synodic month; ν increases by 0·81272 hours in the course of a mean solar day.

The differences between apparent and mean *lunar* time are much greater than those in the solar case. The following extreme cases illustrate the variability in the length of the *apparent* lunar day—the time-interval from one transit to the next. In Berlin, on 1893 December 22–3, this interval was 25^h 8·6^m of mean solar time, but on 1913 August 18–19 it was only 24^h 38·7^m, nearly half an hour less. The semi-diameter of the moon's disk was 16′ 47·4″ in the first case and 14′ 43·0″ in the second case; the *actual* moon's motion in its orbit was, in the first case, faster, because of its closer proximity to the earth, so that the *apparent* motion of the moon in the daily celestial rotation appeared slower. On the other hand, the length of the apparent solar day never deviates from that of the mean solar day by more than half a minute of time.

On 1930 January 13–15 the upper transits of the moon followed each other after an interval of 25^h 7·3^m, the moon's semi-diameter had the high value 16′ 47·4″, as in the former example. About a quarter of a month before or after such a day on which the moon approaches the earth so closely, the greatest differences between apparent and mean lunar time are to be expected. In fact, on 1930 January 9 the apparent moon culminated more than half an hour *before* the mean moon, but then speeded up in its orbital motion and, on January 21, culminated half an hour after the mean moon.

In spite of these irregularities, lunar geophysical reductions (other than those relative to the sea-tides) have, in the past, usually been based on the movement of the actual moon—largely because of the convenience of the data furnished day by day for the actual moon in the astronomical year-books. The tidal action of the moon, however, has been expressed, by the harmonic analysis of the tidal forces, in terms of the motion of the mean moon. Therefore it is preferable to use mean lunar time throughout. The geophysical lunar almanac by Bartels and Fanselau [30] gives complete tables for the mean moon for the years 1850–1975. Instead of ν, these tables give (in hours)

$$\mu = 24 - \nu, \tag{1 b}$$

a quantity introduced by Schmidt [32] for the lunar reduction of the Potsdam magnetic data. In addition, the lunar almanac gives data for
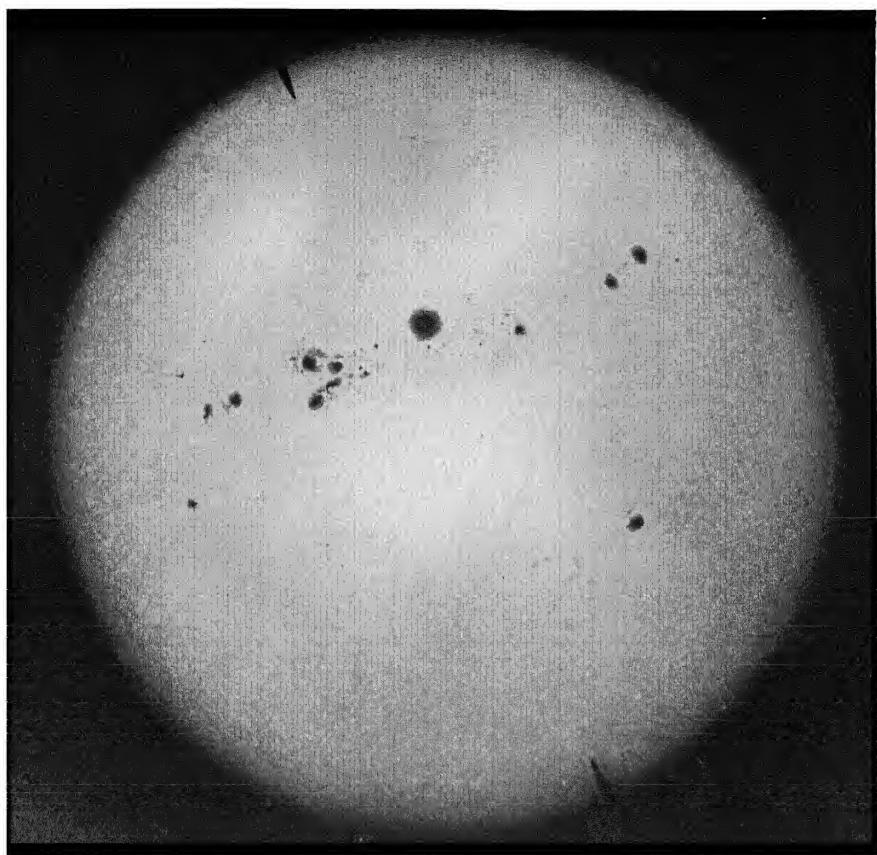
PLATE 15

Photograph of the sun, taken on 1929 November 30 (Mount Wilson Observatory)

PLATE 16
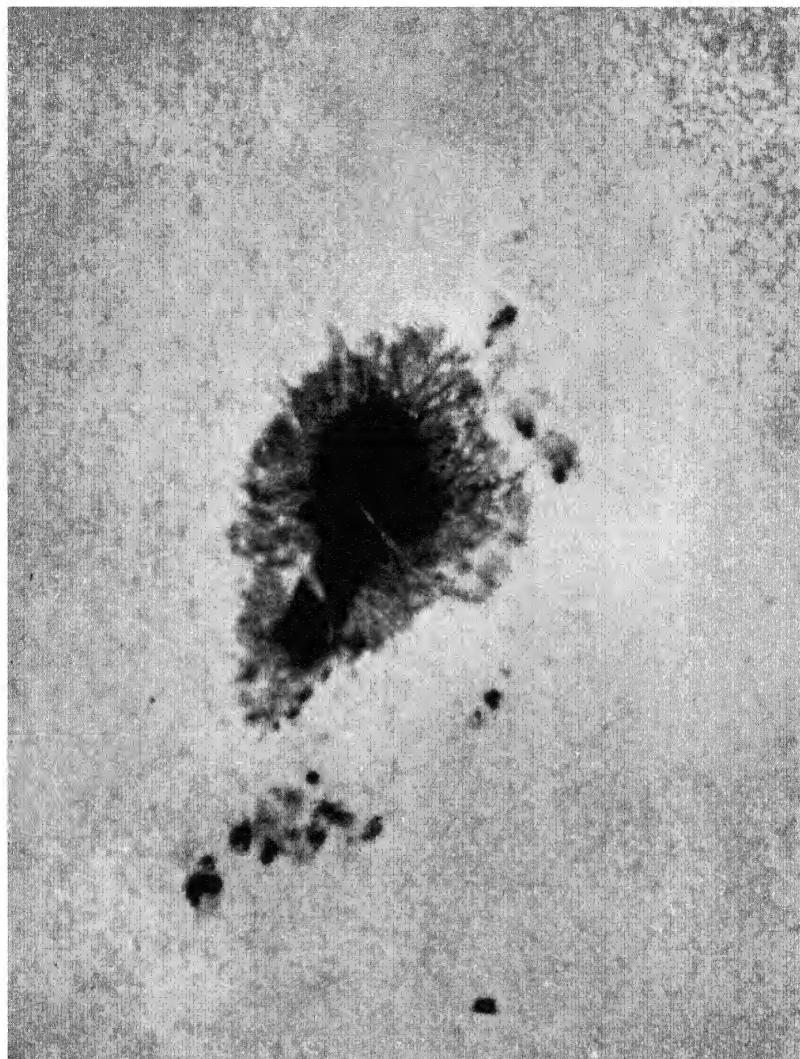
The great sunspot group of 1928 June 30 (Yerkes Observatory)

the mean perigee and apogee of the moon, and for the passage of the mean moon through its ascending node.

5.3. Phenomena on the sun's surface [1]. The sun, like the earth and moon, has surface irregularities, but these are much less permanent than those on the earth. The most obvious of these features on the sun are the relatively dark patches called *sunspots*; these vary greatly in size, but may have diameters of 50,000 km. or even more (Plates 15, 16). Probably the largest on record was one seen in 1858, which was 230,000 km. across, or eighteen times the diameter of the earth. (The diameter of the sun is 1·390 million km.) When their diameter is 40,000 km. or more, they are visible to the naked eye; their motion across the solar disk indicates the rotation of the sun, which is in the same direction as that of the earth.

Sunspots are not of long continuance. Small spots last only for a few days, large ones for a few weeks. Only a few are visible during more than two rotations, but some continue throughout four or even five rotations (of about 27 days each). They differ from one another in form and behaviour as well as in duration, but there is much similarity in the life-history of different spots. Usually spots occur in pairs, though at certain stages in their evolution there are also smaller companion spots. The leading, or western, spot is generally the larger, more symmetrical, and more stable, of the two main spots of a group. (References to east and west on the sun denote the directions as seen by an observer *on the earth*; that is to say, if he is on the earth's northern hemisphere, and so faces southwards when he views the sun at noon, he regards the left half of the sun's disk as the eastern, and the other as the western half; points on the visible hemisphere of the sun move from east to west, and spots on the other hemisphere first come into view at the eastern edge or 'limb'.) Each completely developed spot consists of a central dark portion, called the umbra, with a lighter border round it, called the penumbra; even the umbra is only relatively dark, being at a temperature of about 3,000°—instead of 6,000°, the normal temperature of the *photosphere* (the layer from which most of the sunlight is emitted).

Besides the dark spots, there are intensely bright streaks on the sun, thousands of kilometres long, called *faculae*; they are often situated near sunspots, though they are only to be seen easily when near the edge or 'limb' of the sun, which is darker than the centre of the disk. They are of branching form, and often change their shape rapidly in a few hours—though their general position on the sun may

be nearly constant for weeks. They lie above the level of the photosphere.

The *spectroheliograph* enables photographs of the sun to be taken by the light of a very narrow band or line in the sun's spectrum; this light is produced mainly by a single constituent of the atmosphere, whose distribution can thus be found. The Fraunhofer H and K lines of singly ionized calcium, in the violet (wave-lengths 3,968·5 and 3,933·7 A, where A = Ångström unit = 10^{-8} cm.), are specially convenient for photography, but the red H_α hydrogen line (6,563 A) is also used (Plate 17). Different parts of the calcium lines (which are of complex structure) are produced at different levels in the sun's atmosphere, and this fact makes it possible to determine the distribution of this element (and likewise of others) at different heights (Plates 17, 18).

At the limb of the sun, during eclipses, great flames or *prominences* [33, 34] are often visible (Plates 19–21); by means of the spectroheliograph they can be observed when the sun is not eclipsed, and Lyot, at the Pic du Midi, is able to photograph them in their total light, without an eclipse. There are two main types, called *quiescent* and *eruptive*; the former change their shape and position only slowly, and can be observed for several days on end; they appear to consist mainly of calcium, hydrogen, and helium. Eruptive prominences change rapidly, and show numerous spectral lines of other metals besides calcium. The birth of a sunspot is thought to be accompanied, and generally preceded, by one or more eruptions of prominences, which may partly or completely cover the spot at first, and which often precede it in the direction of the solar rotation. Eruptive prominences may fall back on to the solar surface, or may in whole or in part become detached and travel away from the sun. They are often arched, with both ends on the surface, sometimes at great distances apart (e.g. on 1919 May 29 an arch covered nearly 50° on the limb; it broke up and receded from the sun, being visible to nearly 800,000 km. distance from the surface; this was the largest prominence on record). The velocity of ascent of receding eruptive prominences is found by Pettit [33] to be uniform, but to increase suddenly at intervals to a simple multiple of the velocity before the change, though this last conclusion has been criticized by Waldmeier [36]. The highest velocity observed as yet is 728 km./sec. (1937 Sept. 17) [35]; this prominence passed out of the photographic pictures taken of it, at the record height of 1,000,000 km. (Plate 22). E. A. Milne [14] has shown how selective radiation pressure may repel rising matter from the surface, with velocities increasing up to a limit of
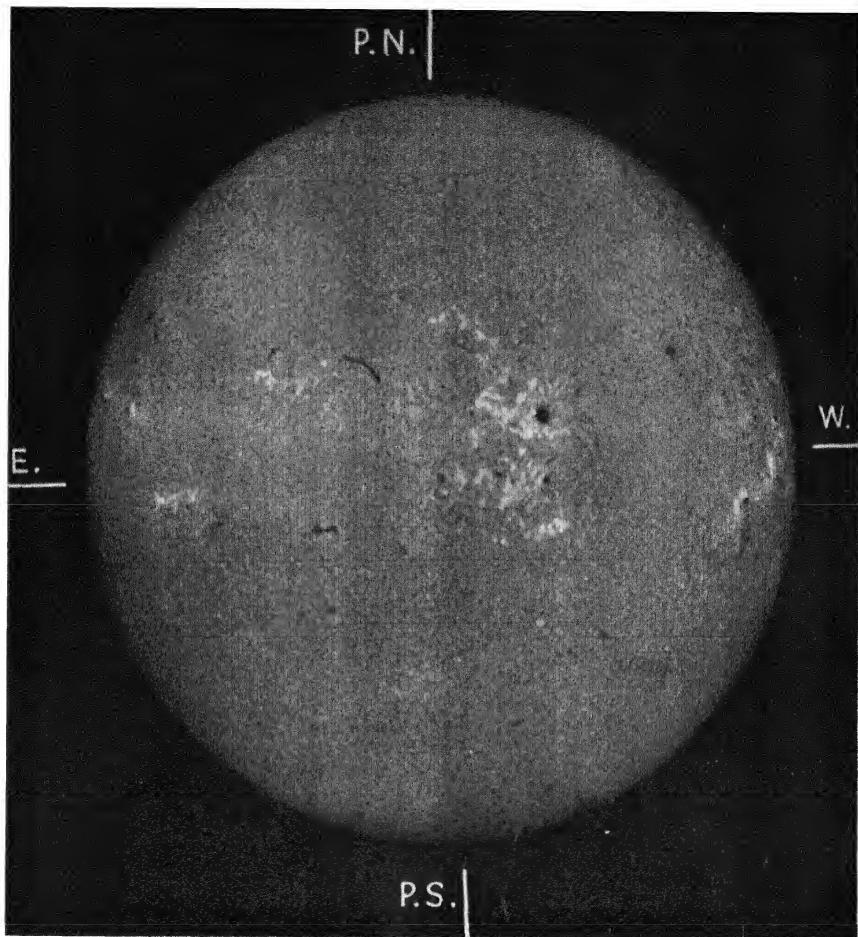
PLATE 17

The sun photographed on 1929 December 17 with the spectroheliograph, in the light of the *H*α line (Meudon Observatory)

PLATE 18
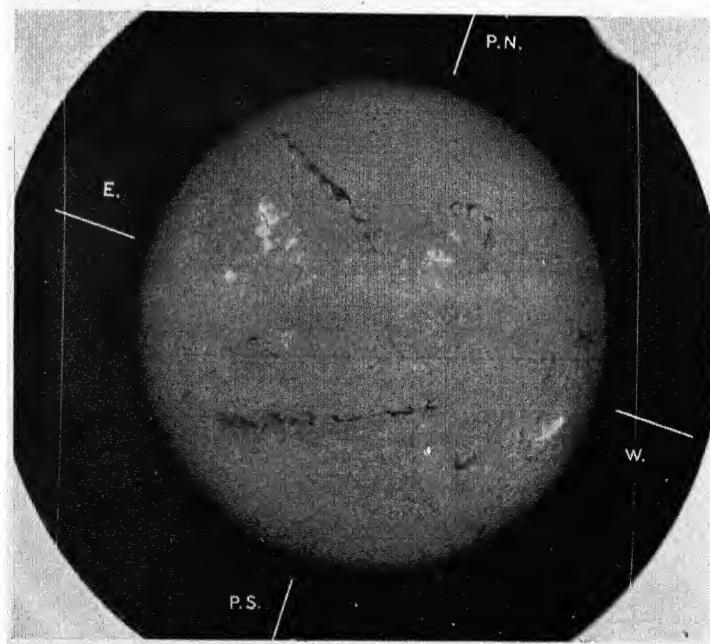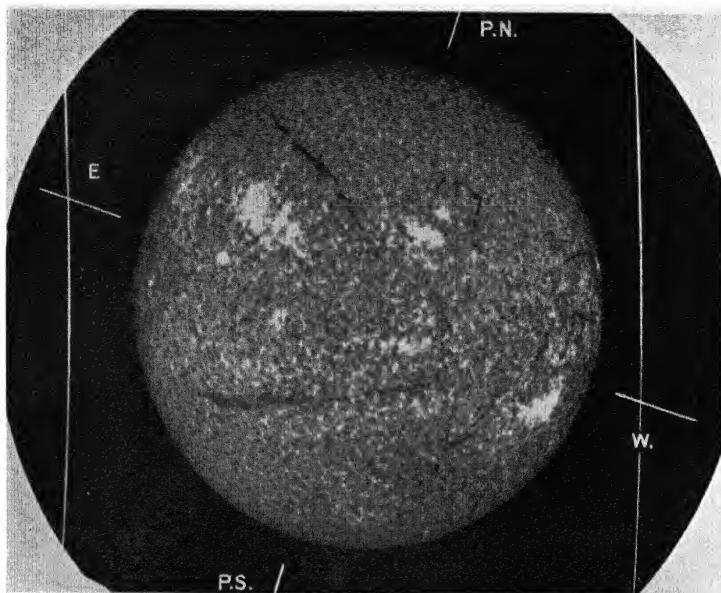
Photographs of the sun, taken on 1930 May 26 with the spectroheliograph, in the light of the K_3 line of calcium (above) and of the $H\alpha_3$ line of hydrogen (below). The light of these lines comes from a high level in the chromosphere
(Meudon Observatory)

PLATE 19
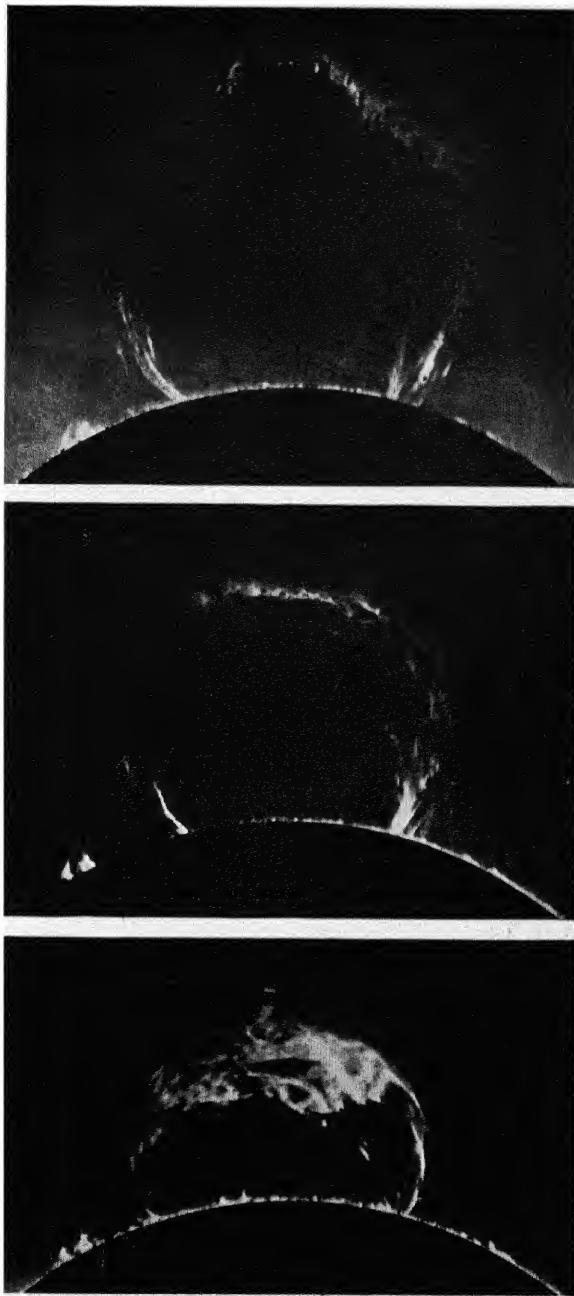
The great eruptive prominence of 1919 July 15, photographed at
the following times (G.M.T.): lower, 3^h 8^m 2^s; middle, 3^h 51^m 56^s;
upper, 4^h 7^m 19^s (Pettit)

PLATE 20
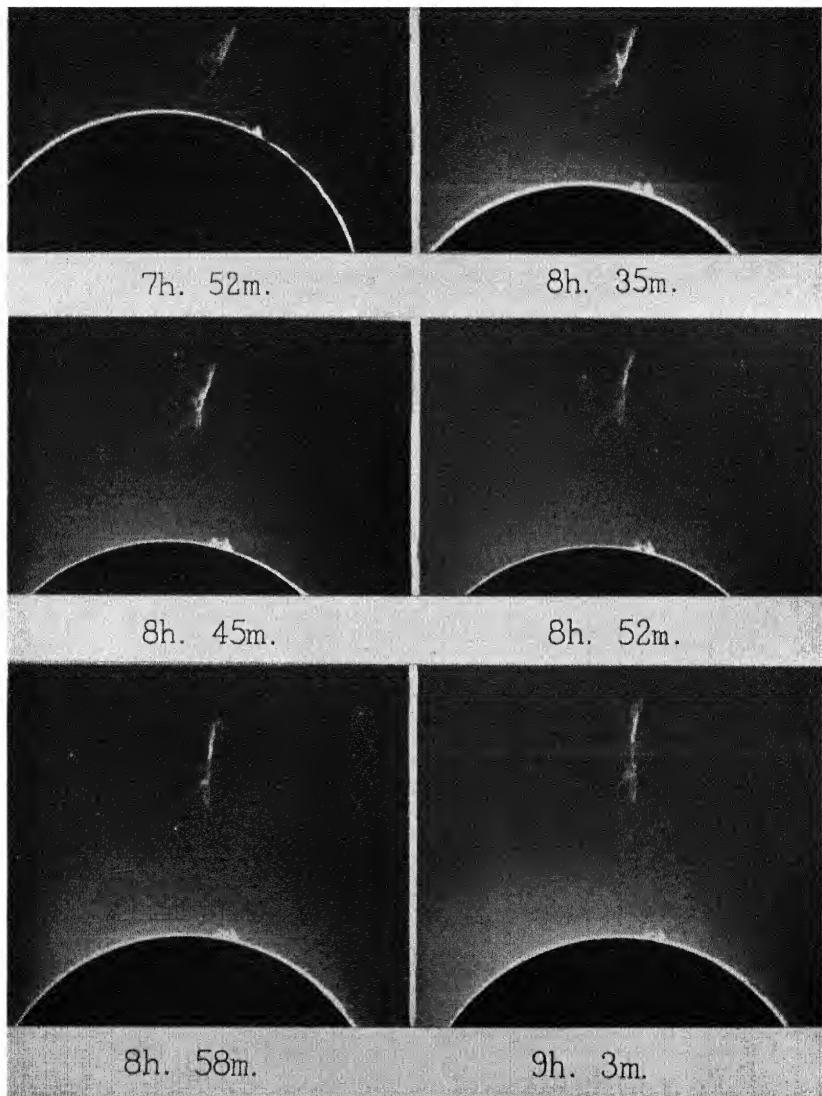
The very high solar prominence observed at Kodaikanal on 1928 November 19 (Royds). The greatest observed height was 20·9′ or nearly one million km.; the greatest observed velocity was 229 km./sec. The times refer to Indian Standard Time

about 1,600 km./sec. if the matter travels right away from the sun; but no satisfactory theory of the observed sudden increases of the prominence velocity has been given.

The spectroheliograph reveals clouds of calcium and of hydrogen on the solar disk, above the photosphere; these are called *flocculi*. The ordinary flocculi change only slowly, but there are other very bright and rapidly changing 'eruptive' flocculi; these always appear in active regions of the surface, and are thought to be identical with the eruptive prominences seen at the *edge* of the disk. Photographs in the light of hydrogen lines reveal dark hydrogen flocculi as well as bright ones; the bright ones are often eruptive and associated with sunspots; the dark ones may be in the cooler, higher regions of the atmosphere (see Plate 18).

The invention of the *spectrohelioscope* [2] by Hale in 1926 has made possible a nearly continuous watch of the sun in monochromatic light, which is organized internationally under the auspices of the International Astronomical Union [1 c]. The number of observed *solar eruptions*, in which patches of the sun's surface increase suddenly in brightness, has therefore increased in recent years. The first observation, by Carrington and Hodgson on 1859 September 1, will be described on pages 333–5.

Newton and Barton [12] characterize solar eruptions, from observations in 1935–6, as follows:

'They appear very suddenly, rise quickly to a maximum, in 5 or 10 minutes, and die away less quickly. An hour after an eruption has first appeared, all traces of it may have disappeared. They last about 20 to 40 minutes, the duration increasing with intensity. In central intensity they sometimes equal or exceed that of the continuous spectrum adjacent to the line in the spectrum in which they are being observed. They are nearly always associated with sunspots, but the photosphere, including the structure of sunspots in the vicinity, appears unaffected.

'The eruptions do not appear to be areas of high temperature black-body radiation. Occasionally two or more bright eruptions occur nearly simultaneously in widely separated groups of sunspots, and such occurrences are more frequent than can be ascribed to pure chance.

'The areas of the brilliant monochromatic emission range from tiny patches to large patches or streaks, between, say, 20 to 500 millionths of the sun's hemisphere.

'The brilliant patches, which themselves usually possess no measurable outward velocity, are often closely associated with absorption markings possessing high radial velocity up to 500 km./sec. Outward velocities are often

followed by inward velocities indicating that part, at least, of the ejected cloud of gas is falling backwards to the sun.'

This description suggests that a bright eruption, seen on the sun's disk, would appear as an eruptive prominence if seen near the limb.

The numbers of bright eruptions, classified according to their observed intensity on an approximate scale 1 (small), 2 (medium), 3 (largest and most intense), in three years of increasing solar activity, were as follows:

Year	Number of bright eruptions of intensity			Total	Mean sunspot number
	1	2	3		
1934	14	1	0	15	8·7
1935	137	18	2	157	36·1
1936	472	134	27	633	79·6

Our knowledge of the changes in the form and position of solar prominences and other solar features has lately been greatly increased by the photographic motion-pictures made by Lyot [37, 38] at the Pic du Midi using the total light (the sun's disk being blocked out, and careful precautions being taken to reduce scattered light in the instrument, that of the sky being already much reduced by reason of the high altitude of the observatory), and by McMath and his colleagues at the McMath-Hulbert Observatory, Michigan, using the light of individual spectral lines [34, 35].

Extending far outwards from the sun there is the rarified envelope called the *corona* [9], formerly observable only during eclipses (Plates 23-4), though Lyot is now able to photograph it regularly at the Pic du Midi [37, 38]. Its form undergoes striking changes, which are associated with the varying frequency of sunspots. The form of the corona seems often to be related to that of the prominences that may exist near the sun's surface at the time. Near the poles of the sun there are coronal streamers or plumes which in shape suggest the lines of force of a spherical magnet; in low solar latitudes long coronal streamers appear, which have sometimes been observed to extend outwards for a distance of many solar diameters. (See also p. 193.)

5.4. The sun's rotation. The motion of spots and faculae across the disk, from east to west, shows that though the sun rotates, the rotation is not like that of a rigid body; the angular velocity at the surface diminishes from the equator towards the poles. Individual spots or spot groups also show proper motions (as well as internal relative

PLATE 21
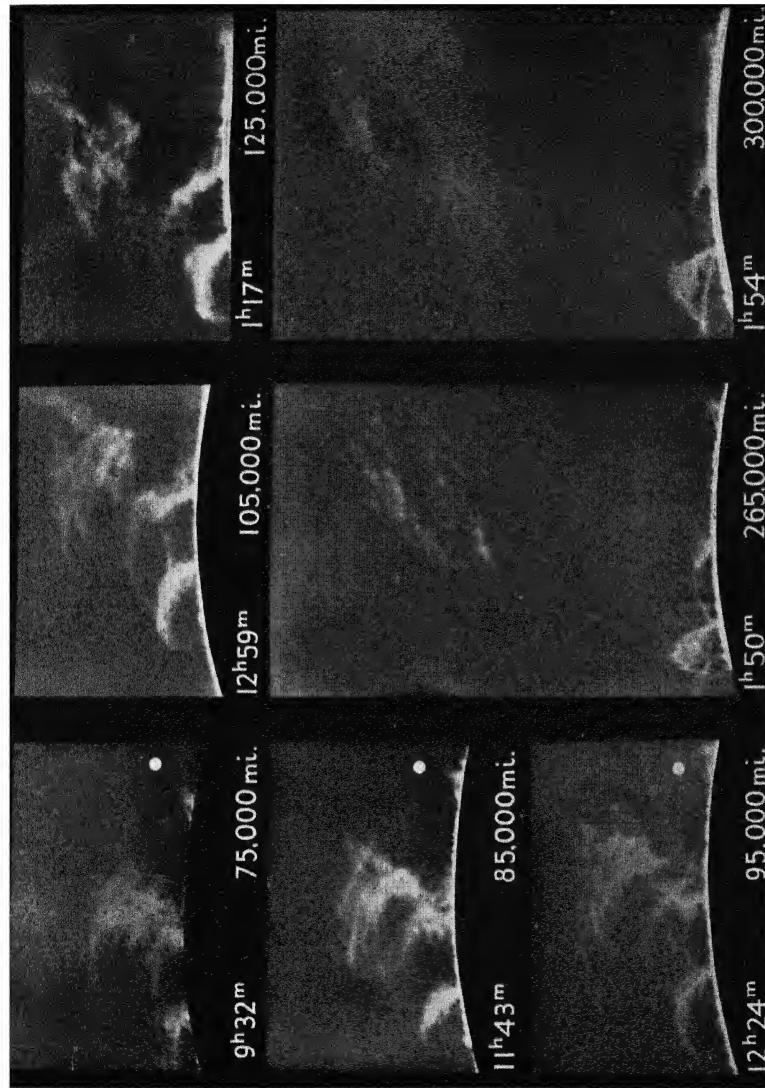
An eruptive prominence which was observed to rise to a height of 300,000 miles, photographed in the light of the K line of calcium, on 1920 October 8 (Yerkes Observatory). The comparative size of the earth is shown by the white disks in the three photographs on the left

PLATE 22
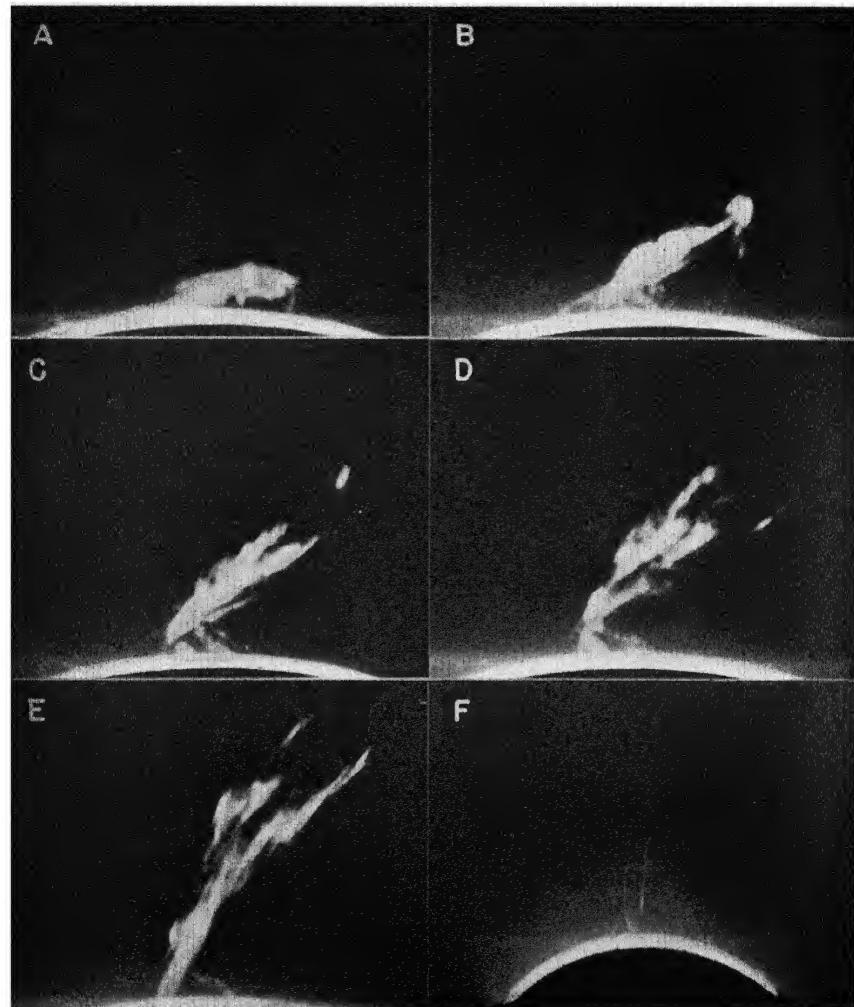
The great eruptive prominence of 1937 September 17, photographed at the McMath-Hulbert Observatory, at the following times (G.C.T.): A, 14^h 50.69^m; B, 14^h 55.84^m; C, 15^h 06.13^m; D, 15^h 09.11^m; E, 15^h 14.31^m; F, 16^h 06.7^m

The exposures A to E were made with a 20-foot focus mirror; F was made with a lens of 74 inches focal length. In F the upper part of the prominence goes out of the picture, at a height of a million miles, with a velocity of 728 km./sec.; the lower part of the prominence is falling into the chromosphere
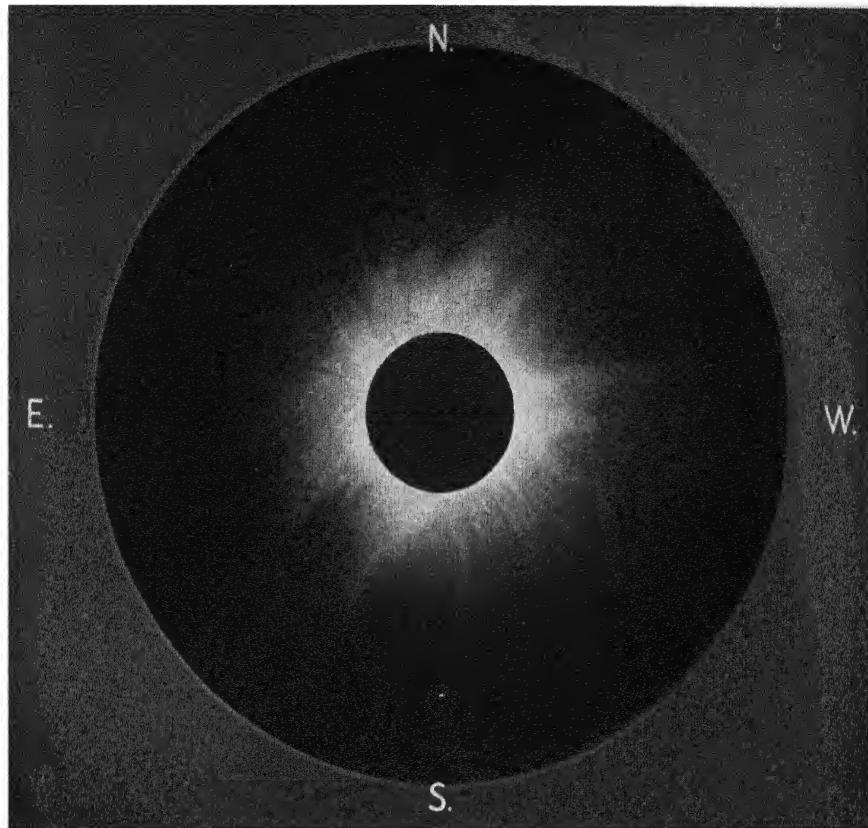
PLATE 23

A drawing of the solar corona on 1905 August 30, from photographs taken at Sfax, Tunisia. (After F. W. Dyson.) The epoch was near the sunspot-maximum of 1906·0

PLATE 24
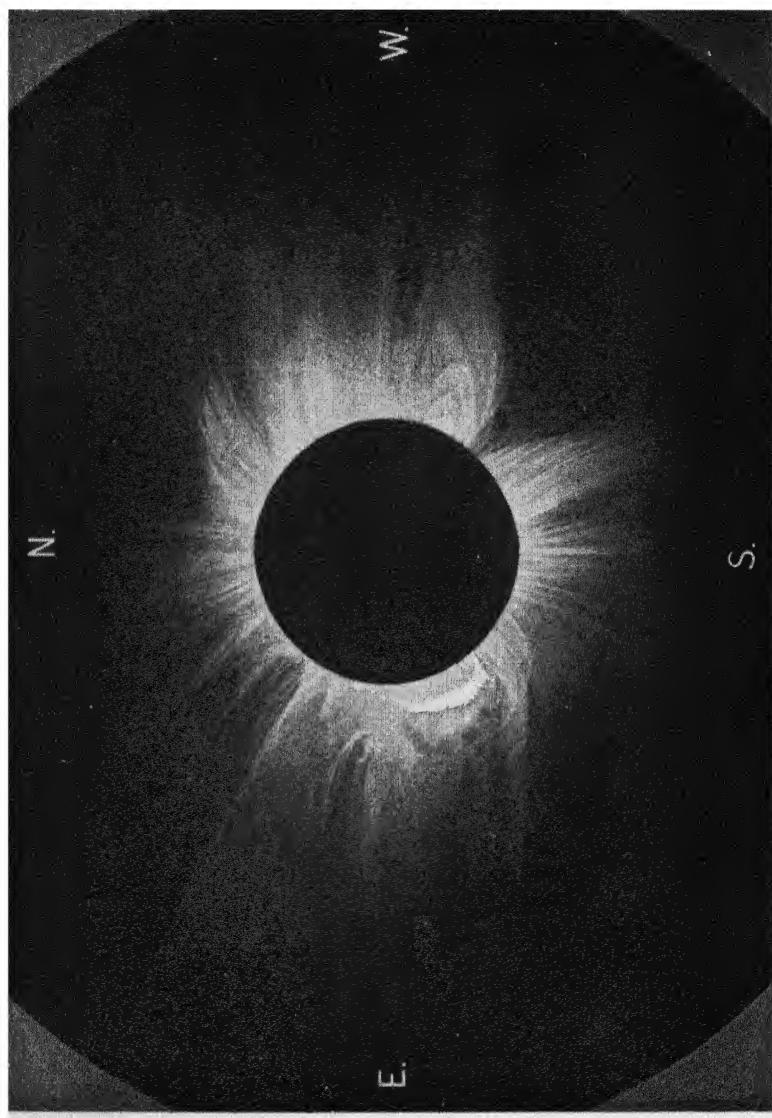
A drawing of the solar corona on 1919 May 29, from photographs taken at Sobral, Brazil. (After F. W. Dyson.) The epoch was between the sunspot-maximum of 1917·5 and the minimum of 1923·5

motions) upon the surface, which result in small changes of latitude, and differences between the rotation periods obtained from different spots for the same latitude.

The mean sidereal rotation period found by Carrington from sunspots in all latitudes where they were observed was 25·38 days; this corresponds to a mean *synodic* rotation period (i.e. the period of rotation relative to the earth, or the time between successive passages through the central meridian) of 27·275 days, though on account of the earth's varying orbital motion the actual synodic rotation period varies slightly throughout the year. In order to specify the position in longitude of any point upon the sun's surface at any time, the following procedure has been adopted at Greenwich, where, in co-operation with other observatories, the positions and areas of sunspots have been measured daily almost without a break since 1874. A meridian rotating uniformly with the sidereal period 25·38 days, and which passed through the ascending node at the epoch 1854·0, was taken by Carrington as the *prime meridian* from which *heliographic longitude* is measured. The celestial longitude of the ascending node (the point B_1 in Fig. 3) was taken as 73° 40′ for 1850·0. The successive *rotations* of the prime meridian are numbered, the first being that which commenced on 1853 November 9.

The adopted sidereal period corresponds to that of sunspots in heliographic latitudes 20° or 22° [7 a]; it is somewhat longer than the mean rotation period of all sunspots. The heliographic longitude of spots in lower latitudes continually increases, and for spots in higher latitudes it steadily diminishes. The following Table gives the mean synodic rotation periods derived from spots in different latitudes.

<div align="center">TABLE 2</div>

Latitude	0°	5°	10°	15°	20°	25°	30°	35°
Synodic period in days	26·4	26·5	26·6	26·8	27·1	27·5	27·8	28·7

The diagram, Fig. 5, illustrates the equatorward increase of angular velocity of sunspots: it is drawn on the assumption that there is a spot in every fifth degree of latitude, and that at a given moment these spots were all on the sun's central meridian. If, then, each spot travelled with the average speed of apparent motion appropriate to its own particular latitude, in 27·275 days the spots would be found in the positions indicated on the curved line.

It is important to note, however, that the rotation periods determined from different spots in any one narrow belt of latitude cover

a range at least as wide as that of the average periods for each latitude shown in Table 2.

Table 3, due to Mr. and Mrs. Maunder [7 a], illustrates this; it refers to spots observed at Greenwich during the 23 years 1879–1901, including almost all those that lasted for 6 or more days—about 1,870
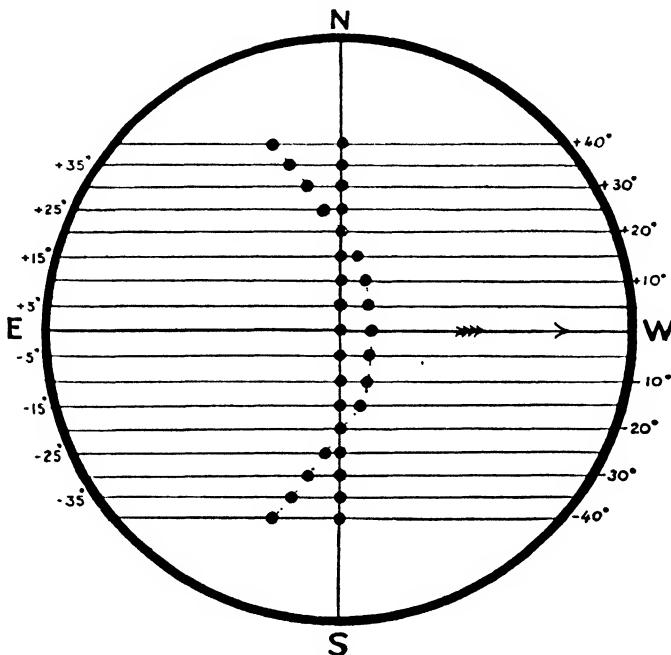
FIG. 5. The relative apparent movements in longitude of sunspots in different solar latitudes during one 'standard' rotation of the sun.
(After E. W. Maunder)

in all. Their mean speeds across the disk were expressed in terms of the corresponding periods of complete rotation. The table shows the number of spots in each band of latitude that had synodic rotation periods within ± 0.1 day of the values given in the first column. The extreme range is from 24·4 to 31·2 days, or 6·8 days, whereas in Table 2 it is only 2·3 days.

Mr. and Mrs. Maunder found that the spots of long duration, which have more than one passage across the sun's disk, give a somewhat greater mean period than the short-lived spots, and are more accordant *inter se*.

Besides the motion of a spot group as a whole, internal movements are shown. In the early days of a group the leading spot (the western spot, as reckoned by an observer on the earth) tends to rush forward

TABLE 3

Synodic rotation period d.	Below −35°	−30° to −35°	−25° to −30°	−20° to −25°	−15° to −20°	−10° to −15°	−5° to −10°	0° to −5°	+5° to 0°	+10° to +5°	+15° to +10°	+20° to +15°	+25° to +20°	+30° to +25°	+35° to +30°	Above +35°
23.4					2	1										
23.6					1											
23.8					1	2										
24.0			1	1			2				1					
24.2					2	2	1				3					
24.4					6	1	3			1	2	2				
24.6			1	1	1	3	1			1	2		3			
24.8			1		3	3	4	2			5		1			
25.0				2	8	6	6	1	1	3	3	2	1			
25.2			1	2	9	3	4	2	1	2	5	3	3	1		
25.4			1	8	9	13	5	1	2	4	12	5	1			
25.6				7	16	15	5	3	1	2	12	4	4	1		
25.8			2	4	14	13	12	2	4	7	12	9	4	1		
26.0			4	9	14	11	15	1	2	15	23	6	5			
26.2			1	7	16	16	26	3	1	10	21	7	8	2		
26.4			2	8	23	24	19	2	1	16	23	15	9	1		
26.6			1	12	38	26	20	2	10	14	31	15	3	3		
26.8		1	4	15	33	30	42	15	10	7	41	14	7	2		
27.0			4	14	39	41	26	6	10	26	39	25	13			
27.2	1		6	13	16	37	19	15	9	17	29	25	15	1	1	
27.4		1	1	8	21	23	5	7	9	15	7	18	12	2		
27.6			3	7	4	13	5	6	8	8	9	22	11	1		
27.8		1	13	5	3	4	5	5			4	5	5	3		
28.0					6	4	2	1	1	3	2	4	5	4		
28.2		1	1	1	1	3	2	2		1	1	3	5	4		
28.4		1	1		1	4	1	1		2	2	2	1	2	2	
28.6		1	1		1	2		1					2	1		1
28.8		1		1		1	1	2			1			3		
29.0			1		1			1					1	1		
29.2					1	1					1			1		
29.4		1											2	1		
29.6					1	1					1					
29.8											1			1		
30.0		1														
30.2											1					
30.4																
30.6																
30.8																
31.0																
31.2																

over the surface; in the group's later history it slackens speed and returns on its track. Also the groups which live the longest, returning two or more times to the visible hemisphere of the sun, move more slowly on the whole than short-lived spots, and yield different rotation periods during their successive appearances. ,This is illustrated in Fig. 6, which exhibits the distribution in longitude of all groups lasting through
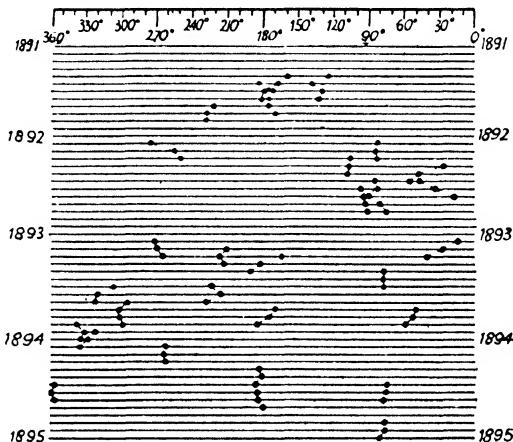
FIG. 6. The heliographic-longitude distribution of sunspot groups of long duration, observed during the years 1891–4. (After E. W. Maunder)

three or more successive rotations during the years 1891–4; for spots that rotate with the same period as the prime meridian, the heliographic longitude when on the central meridian of the sun (i.e. the meridian through the earth's centre) is the same in successive rotations. Hence when the dots in the diagram relating to a given group are not vertically in line the motion of the group indicates a rotation period differing from the mean; when they are not collinear, the group has rotated with different speeds in the successive rotations.

Higher levels of the sun's atmosphere rotate faster than the sunspots. From a table of angular velocities given by Abetti [1], the following synodic rotation periods were calculated:

TABLE 4

Heliographic latitude	Sunspots	Reversing layer	Mean level of emission of the calcium line 4,227 A	Mean level of emission of the hydrogen α line (near the limb)	
0°	26·9	26·4	25·9	25·7	days
15°	27·3	27·0	26·0	25·9	days
30°	28·3	28·4	27·0	26·4	days

5.5. The 11-year sunspot cycle. The number of spots visible on the sun depends partly on the solar rotation, owing to which no spot can remain visible for more than about 14 days on end. Apart from this, however, the average number of visible spots varies greatly from year to year, but with a certain rough regularity of increase and decrease between years of sunspot-maximum and sunspot-minimum. Each complete alternation of increase and decrease, from one minimum to another, is called a *sunspot cycle*. The duration of sunspot cycles, in years, varies from 10 to 13, with an average value 11·1. The degree of sunspottedness at the maximum epoch also varies from one cycle to another.

Figs. 7 and 8 illustrate different aspects of the variation. The former shows the annual mean spotted area for the period 1854–1912, indicating sunspot-maximum years in 1860, 1870, 1883, 1893, and 1905. The average spotted area in 1870 was 1·5 thousandths of the sun's visible hemisphere. Fig. 8, giving [7] the annual percentage of spotless days for the period 1826–1913, shows that in this respect the sunspot-minimum years stand out as definitely as the sunspot-maximum years in respect of spotted area.

In cycles of high sunspot-maximum the rise from minimum to maximum is much more rapid than the subsequent return to minimum; the two intervals of increase and decrease are about four and seven years. In 'weak' cycles with a low maximum sunspottedness the rise and decline are about equally rapid; Fig. 9 illustrates this, after Schmidt [8 a], in terms of the *relative sunspot numbers* (*v. infra*), for the seven strong cycles beginning in 1766, 1775, 1784, 1833, 1843, 1856, 1867, and for the seven weak cycles beginning in 1755, 1798, 1810, 1823, 1878, 1889, 1901.

Turner [8] compared the sunspot curves for alternate 11½-year cycles and found a certain average difference, shown in Fig. 10; the 'first-half' curve refers to the cycles beginning in 1842, 1865, 1888, and the other curve to the cycles beginning in 1853, 1876, 1899. Schmidt, however, found no difference in the average curve for alternate cycles, from the fourteen cycles extending from 1755 to 1901.

The sunspot areas are published annually by the Royal Observatory, Greenwich, in a volume entitled *Greenwich Photoheliographic Results*; the measures are made from daily photographs of the sun taken at Greenwich and other co-operating observatories; this work continues an investigation due originally to Carrington and de la Rue.

A less precise but very convenient measure of solar spottedness is

Mean Latitude of Spotted Area.
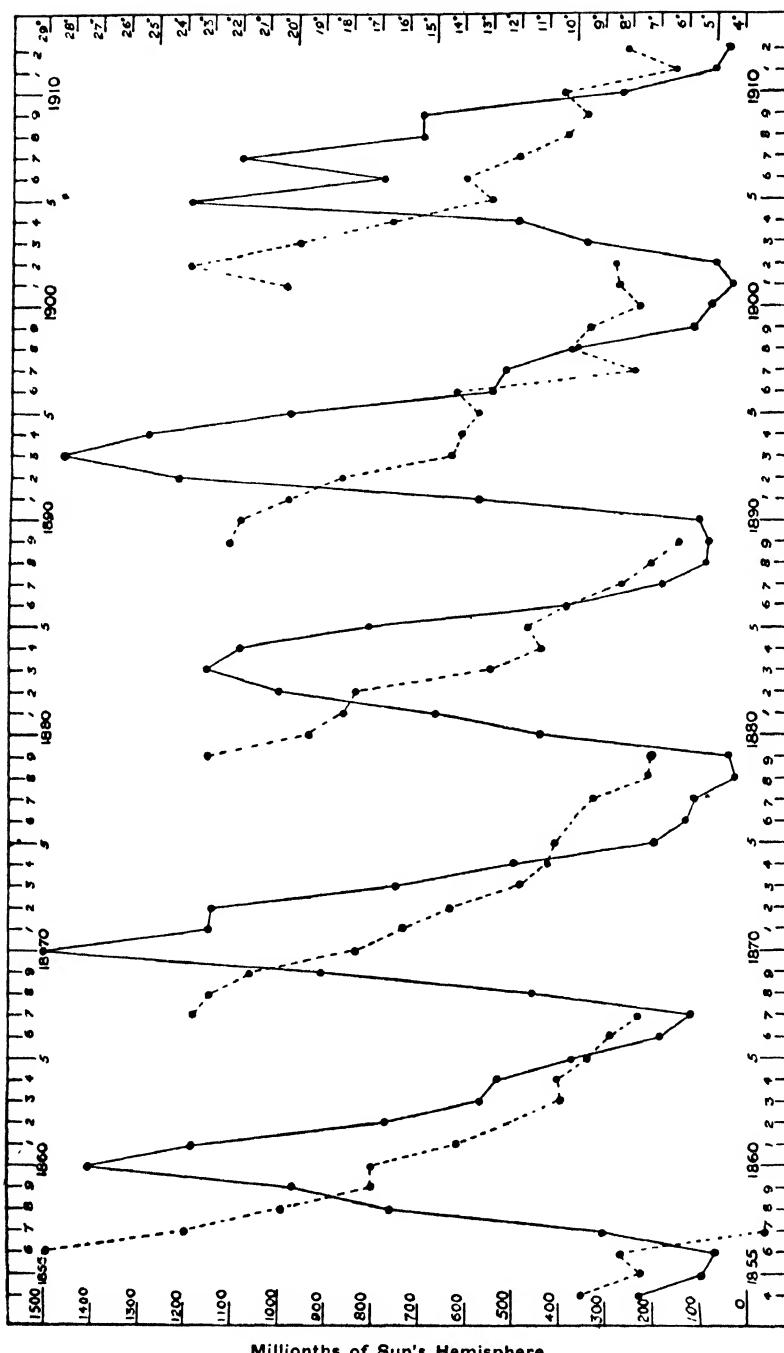
FIG. 7. The mean sunspot area (continuous line) and the mean heliographic latitude of sunspots (broken line), 1854–1912.
(After E. W. Maunder)

Millionths of Sun's Hemisphere.
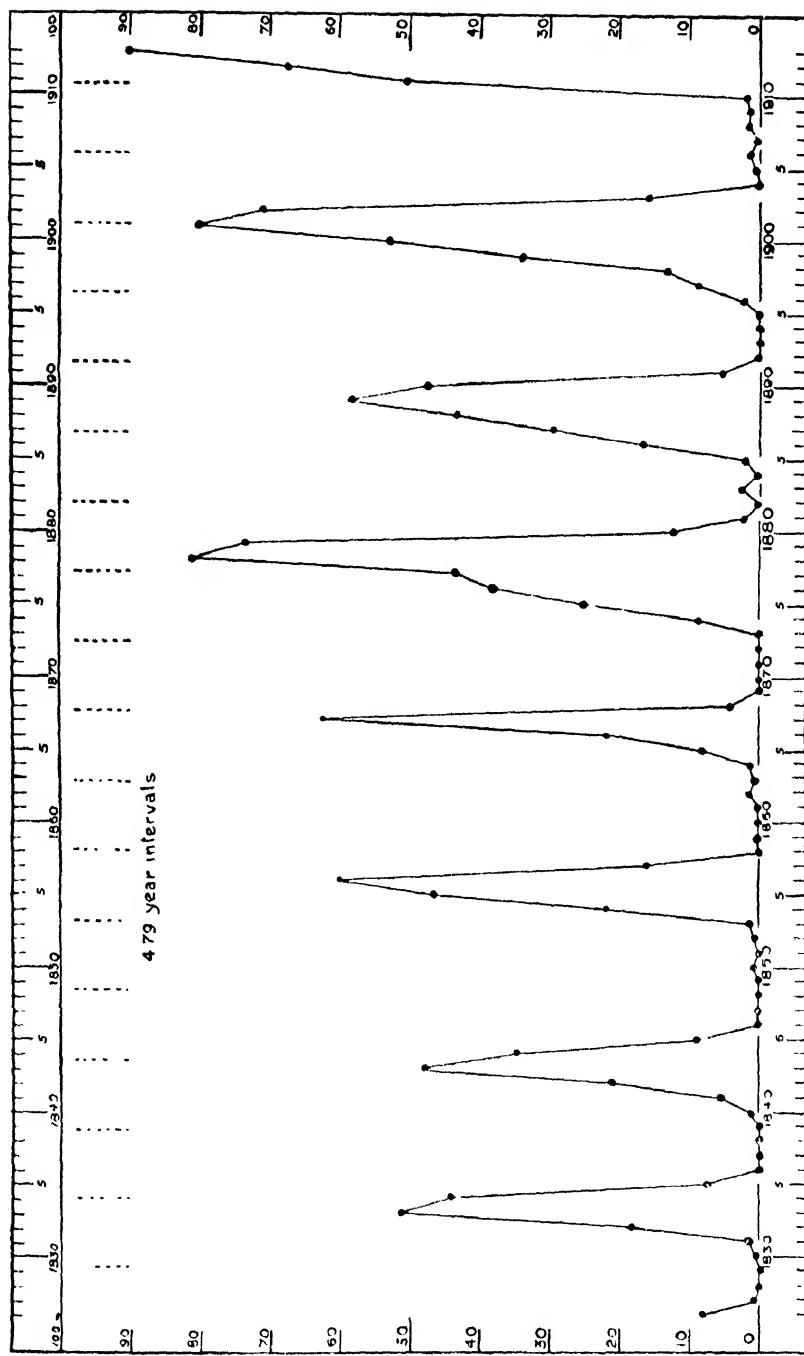
Fig. 8. Annual percentages of spotless days, 1826–1913 (the early observations are due to Schwabe at Dessau). (After E. W. Maunder)

afforded by the *relative sunspot numbers s*, whose preparation was begun by Wolf at Zürich, and has been continued there by Wolfer and Brunner; the number *s* for any day is derived from data contributed by various observatories co-operating with the Zürich observatory; its
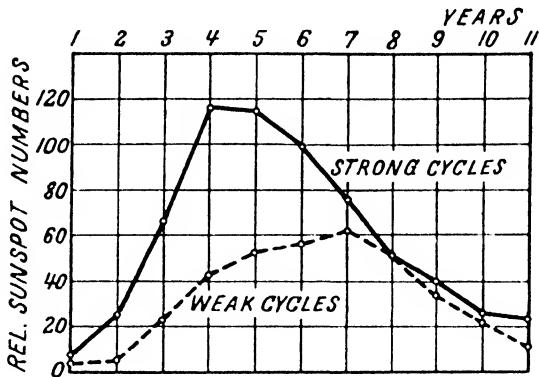
FIG. 9. The average variation of the annual mean sunspot numbers during strong sunspot cycles (continuous line) and during weak cycles (broken line). (After A. Schmidt)
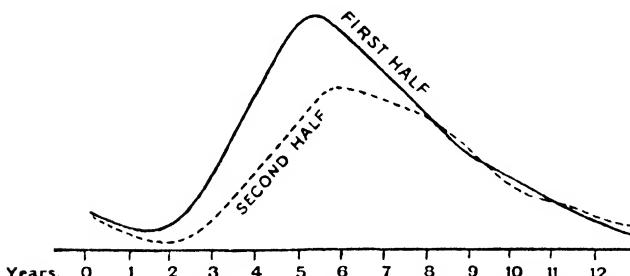
FIG. 10. Alternate halves of the 23-year sunspot cycle compared. (After H. H. Turner)

value is $k(10g+f)$, where f and g denote the number of individual spots, and of spot groups, visible on the sun at the observatory on the date in question, and k is a factor depending on the observer and his instrument, chosen so that the numbers from all the observatories shall be combined on a uniform plan, from year to year. The numbers have been carried back as far as 1749 from old records; monthly mean sunspot numbers for every month from 1749 to 1924 have been published by Wolfer. The daily numbers are published from year to year in the *Astronomische Mitteilungen.* Tables of sunspot numbers are given at the end of this book (Vol. II).

Sunspots usually lie within 30° of the sun's equator. Each cycle begins with new activity at the higher limit of spot latitude; from the beginning of the rise towards maximum sunspottedness, and throughout the subsequent decline to the next minimum of spot numbers and area, the mean latitude of the spots decreases. In the last phase the spots, few and short-lived, are near the equator, where ordinarily spots are rare. Before the last traces of expiring activity have ceased, the earliest spots of the new cycle appear in high latitudes. The new and old cycles may overlap in this way by as much as two years, but on account of this remarkable progress in latitude the two cycles of activity can be clearly distinguished. Taking this into account, the curves of annual mean sunspot latitude in each cycle are discontinuous; they are shown in Fig. 7 by dotted lines [7].

Fig. 11, due to Maunder, shows the same feature in a different way, which also distinguishes between the northern and southern hemispheres of the sun; these are by no means equally affected by spots in any given cycle: in some cycles one, in others the other hemisphere predominates [7]. It indicates the distribution in latitude of the centres (only) of 7,000 spots, entirely without reference to their areas. Each rotation, according to the Greenwich numeration (§ 4), during the interval 1874–1913 covering nearly four cycles, is represented on the diagram by one or more parts of a straight horizontal line, the lines for successive rotations being drawn at equal intervals. Distances along the lines, measured from the dotted centre line of the diagram, correspond to solar latitudes. Whenever, on one or more days during a given rotation, the centre of a group of spots fell in the 1° interval of latitude between $n \pm \frac{1}{2}°$, n being an integer, the corresponding interval along the line for that rotation was filled in. Each sunspot cycle is clearly marked by a 'butterfly' arrangement of the lines, showing the commencement of activity in high latitudes, and the gradual decline of latitude as the cycle proceeds. Every sunspot latitude has its own broad period of activity once only in each cycle. The diagram also suggests that during the cycles that had their maxima in 1883 and 1893 the southern hemisphere had distinctly more spots than the northern; the figures for the spot areas, which in Fig. 12 are shown separately for the two hemispheres, show that this was markedly the case, and that the reverse relation obtained during the two following cycles.

5.6. Regions of long-continuing solar activity. Another important feature of the distribution of sunspots relates to their longitude. Maunder showed that over a period of years certain longitudes, in any
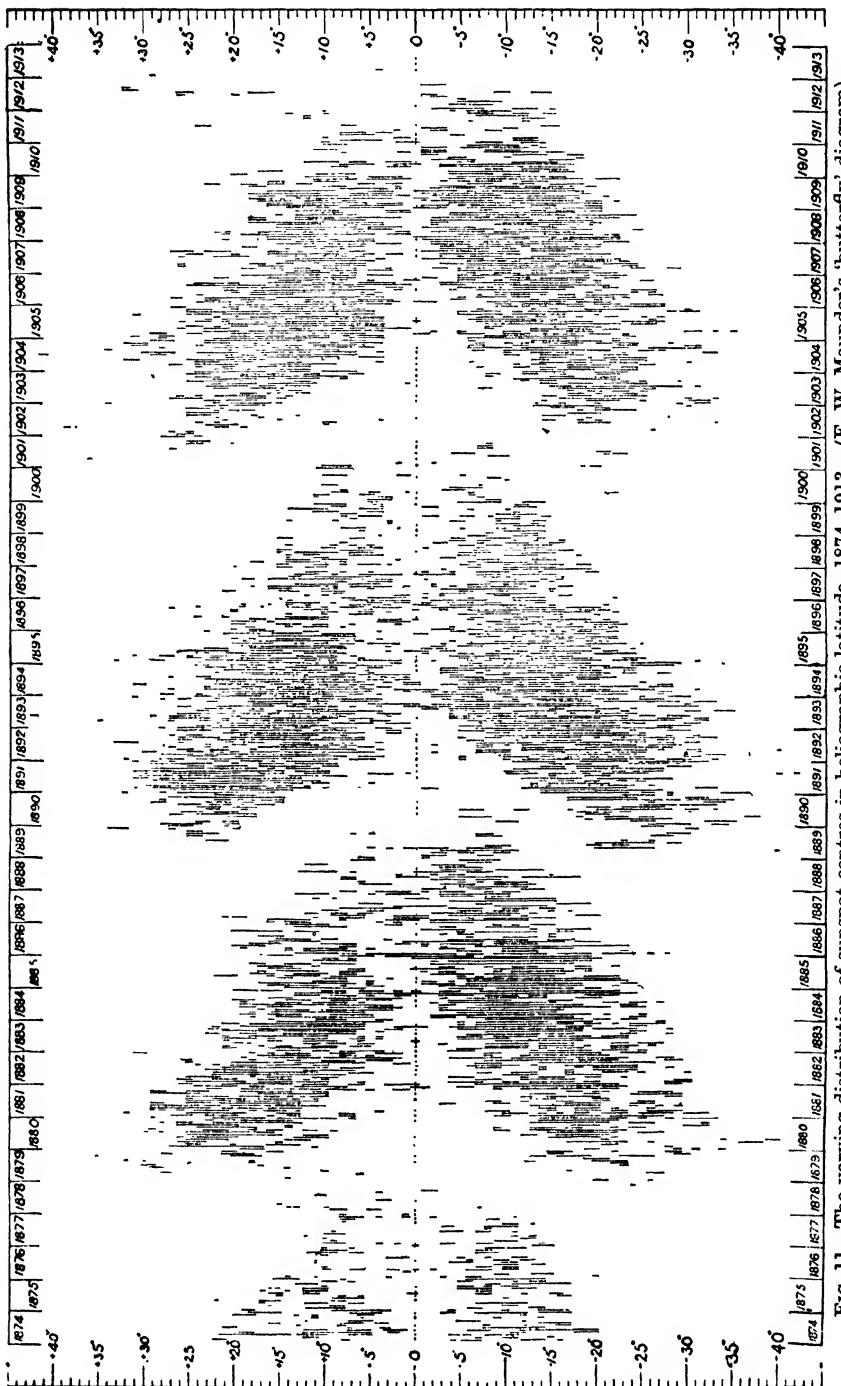
FIG. 11. The varying distribution of sunspot centres in heliographic latitude, 1874–1913. (E. W. Maunder's 'butterfly' diagram).
Scale of heliographic latitude, + to north, − to south
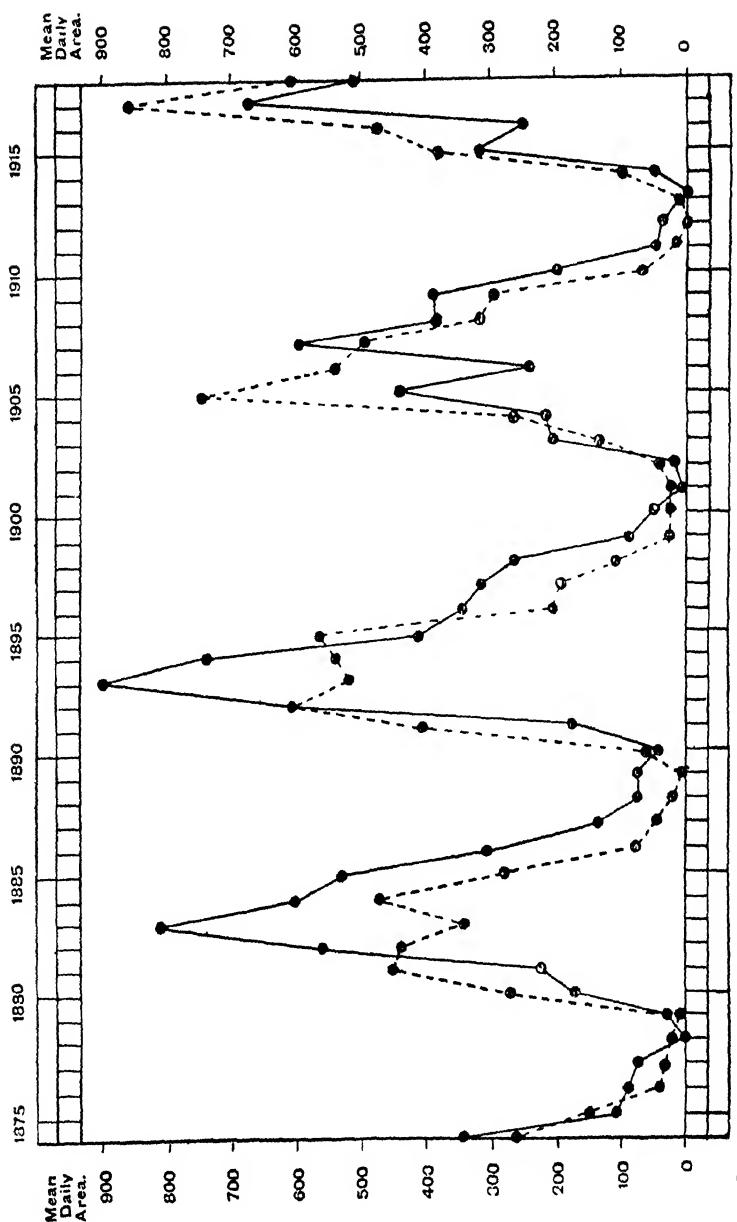
FIG. 12. Mean daily areas of sunspots on the northern hemisphere (dotted line) and on the southern hemisphere (continuous line), 1874–1918. (After E. W. Maunder)

given latitude, tend to be more active in spot-formation than others. This is shown by the following examples [7 a] of spot groups appearing, during periods of about a year, within two small regions (northern and southern) which were intermittently disturbed.

TABLE 5. *Example of an Intermittent Northern Spot Group*

Rotation	No. of group	Mean long.	Mean lat.
534	3191	50·44°	+12·41°
535	3232	53·13°	+11·77°
536	3272	56·27°	+10·63°
Interval of five rotations			
541	3460	53·44°	+9·79°
542	3492	54·85°	+11·90°
Interval of three rotations			
545	3629	56·44°	+7·52°
Interval of four rotations			
549	3777	51·37°	+7·40°
550	3824	52·67°	+10·93°

TABLE 6. *Example of an Intermittent Southern Spot Group*

Rotation	No. of group	Mean long.	Mean lat.
394	985	171·53°	−8·23°
Interval of three rotations			
397	1056	170·19°	−10·21°
398	1071	176·50°	−10·93°
399	1094	177·96°	−10·25°
Interval of two rotations			
401	1130	177·38°	−6·70°
Interval of two rotations			
403	1190·	171·28°	−7·82°
404	1226	167·45°	−9·10°

The same feature is shown in a still more striking and conclusive way [7 a] by Fig. 13, which indicates the longitudes of all the spot groups, of whatever size and duration, given in the Greenwich Heliographic Results as lying between the equator and north latitude 7·5°, for the complete sunspot cycle 1890–1902. The distance of each spot-point from the top of the diagram is proportional to the interval from the commencement of the cycle (taken to be on 1891 March 13·76, Greenwich civil time) to the appearance of the corresponding spot. The first group shown was observed on 1891 August 8, the last on 1901 July 8; the last spot of the preceding cycle (in this belt of latitude) was observed twenty-one rotations earlier, on 1890 January 6, and the first spot of the succeeding cycle, in this zone, on 1904 April 11; thus the diagram shows all the spots in this zone occurring over a period of 14 years and 3 months. When the sunspots of this zone of latitude for the cycle in question were charted, as in Fig. 13, according to their longitudes, but referred to the primary meridian (§ 4) rotating with the synodic period 27·275 days, two narrow but well-marked barren belts were observed to run in a slanting direction across the diagram. This was due to the fact that the sun's equatorial zones rotate faster than this prime meridian. The longitudes of the spot groups were therefore recomputed relative to a meridian with a synodic period of rotation of 26·94 days, appropriate to this zone; Fig. 13 indicates the distribution

SOLAR LONGITUDES

NORTH EQUATORIAL SPOTS, 0° TO + 7·5°
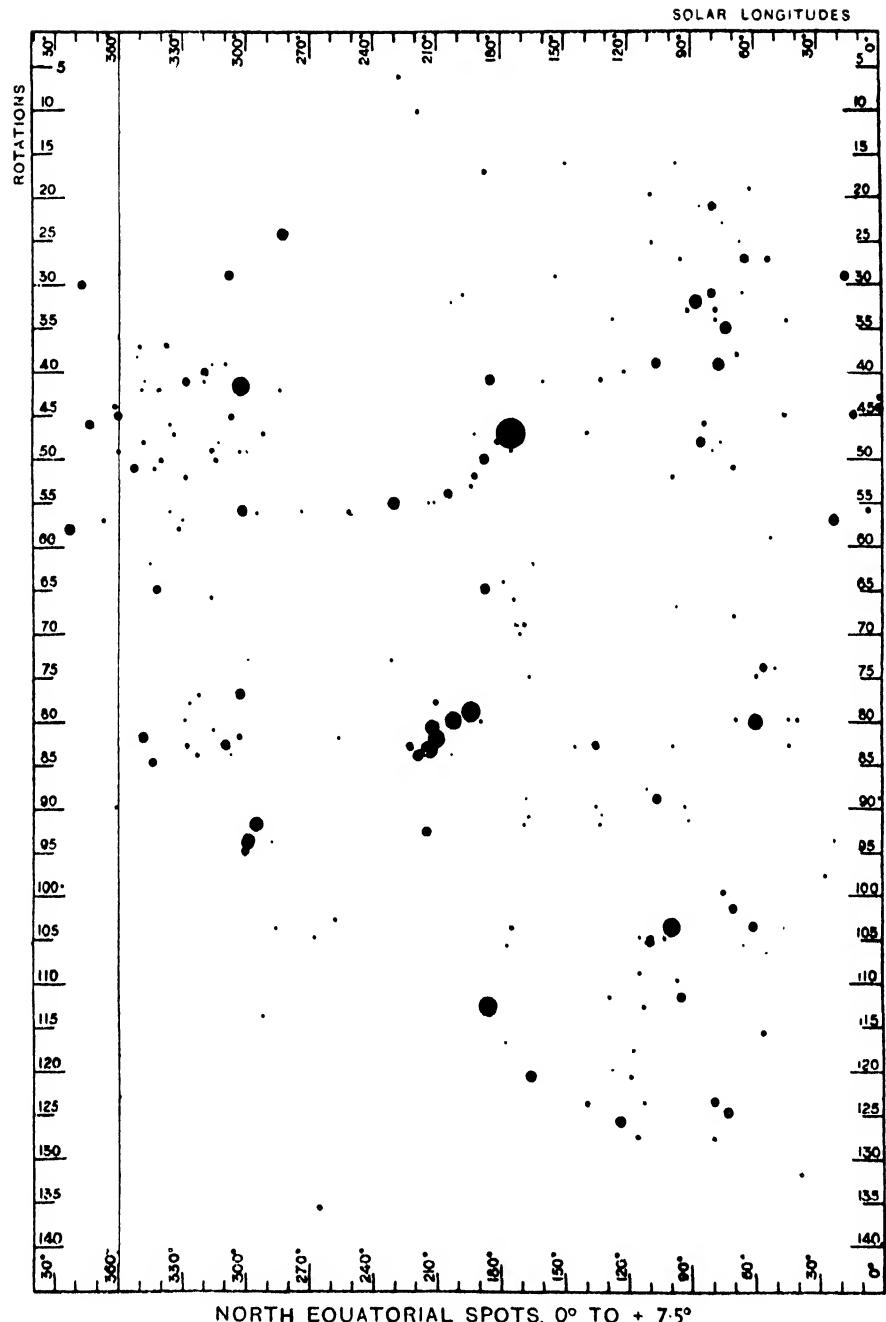
FIG. 13. The solar-longitude distribution of sunspots between the sun's equator and
7·5° N., for the cycle 1891–1902; the sidereal rotation period adopted is 25·09 days, and
the first rotation commences on March 13, 1891. (After E. W. Maunder)

of spots according to these longitudes. The diagram indicates three great regions of activity separated by nearly barren regions; the existence of the barren regions is manifest, in spite of various causes that tend to obliterate it, such as the tendency of great eruptions to spread themselves over long arcs, and the internal motions in the zone (such as the different rotation periods at different latitudes), and even in long-enduring spot groups. The three centres of activity are situated about the longitudes 200°, 340°, and 80°; the barren regions are centred at about 260°, 140°, and 40°, their widths being about 60°, 45°, and 25°. The first great barren belt can be traced, with increasing difficulty, as far as latitude 12°.

Besides showing this differentiation of the sun's activity as regards longitude, the diagram illustrates the intermittency even of the active regions. After a period of growing activity from 1891 to 1895 July 17, for over a year there was an almost complete absence of spots in the zone considered, until 1896 September 18. Such spots as appeared during this period were almost all in a band of only 30° longitude, corresponding to the most active of the three regions. Another epoch of activity followed, affecting all three regions, until 1898 February 23. In the last period the second active region disappeared and the first was much reduced, so that for nearly three years the spot activity was confined almost entirely within 130° out of the 360° of longitude.

5.7. The life-history of large sunspot groups. The first indication of the formation of a group of spots is given by spectroheliograms taken in H or K calcium light, which show bright flocculi in the high solar atmosphere, over a region where the spots appear a day or two later. A typical group is first seen as a pair of small spots almost equidistant from the equator, and about 3° to 4° apart in longitude; the spots grow rapidly and attain their maximum size in about a week, while many small spots appear, usually near the principal components, but also in the space between them. Concurrently the two main spots separate in longitude to 10° or even more apart. The phase of maximum activity lasts a few days, after which the spots begin to decrease; the follower spot is the first to disappear, usually by division into a number of small and diminishing spots, and after a week or so only the leading spot is left. This gradually decreases, without breaking up. The high flocculi persist throughout, and diminish with the surviving leading spot, over which they finally lie symmetrically. The line (or axis) joining the leading (western) and following spots of such a 'bipolar' group is generally only slightly inclined to the sun's equator: the leader

is usually the nearer to the equator; the mean angle varies from almost nothing in low latitudes, to nearly 20° in latitudes 30° or 35°. This is shown by Fig. 14, due to Brunner [11], in which the abscissae denote heliographic latitude and the ordinates the mean inclination of the axes of bipolar spot groups; it is based on Zürich sunspot data for 1894–1928. The 'scatter' of the inclinations about their mean value is considerable;
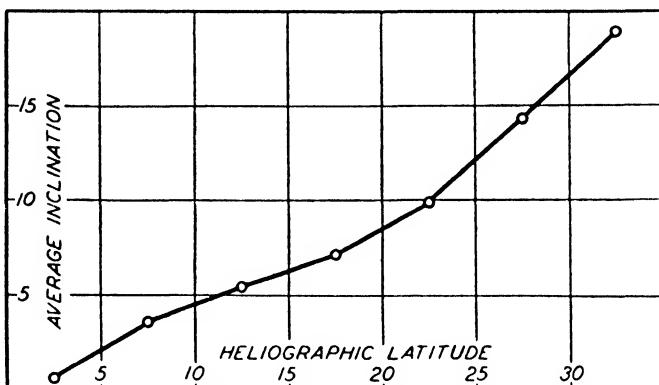
FIG. 14. The mean inclination of the axes of bipolar sunspot groups, to the solar equator, in different heliographic latitudes. (After W. Brunner, from 1,981 groups, 1894–1928)

for example, for the fourth, fifth, and sixth points in the figure the average deviation is 7°, 8·5°, 11°. During the life-period of bipolar spots the mean inclination decreases on the average by 3·2°; the mean inclination for all the bipolar spots is 6·5°.

5.8. Sunspot magnetic fields. Hale, by means of the Zeeman effect, found that sunspots are the seat of intense magnetic fields. The intensities increase with the size of the spot (Fig. 15) and are of the order 1,000 to 4,000 gauss. The two spots in a pair are of opposite polarity.

At the centre of a spot the magnetic force is nearly normal to the sun's surface; the lines of force issue from one spot of a pair and pass (at least partly) into the other spot. The intensity of the force decreases quickly with increasing height in the sun's atmosphere.

The polarities of the leading spots are usually the same during any 11-year sunspot cycle for all spots on the same side of the solar equator, and are of opposite polarities on opposite sides of the equator. At the commencement of each new cycle the distribution of polarities as between leading and following spots, and northern and southern hemispheres, is reversed. This remarkable discovery shows that the

complete cycle of solar changes is about 22 years, and comprises two sunspot cycles.

Fig. 16 shows how definitely the equator divides the sets of spots of opposite polarity; it relates to two bipolar groups observed at Mount Wilson in 1919 August and September, originally in 6° N. and 3° S. latitude. The leading northern spot drifted towards the equator at the
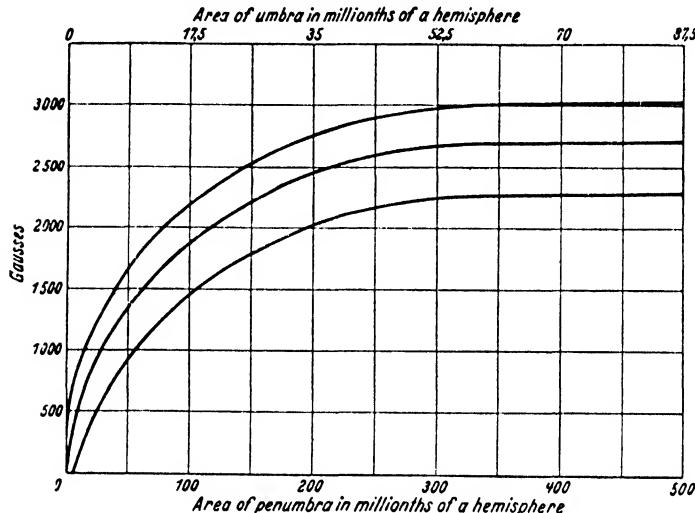
FIG. 15. The intensity of the magnetic field of a sunspot shown as a function of its area; the centre line shows the average relation; the two outer lines indicate the mean deviation of individual spots from the average value for their size. (After S. B. Nicholson, from 1,000 observations at Mt. Wilson, 1917–20, on unipolar spots or on the preceding members of bipolar groups)

rate of 7,200 km. per day, while the leading southern spot increased rapidly in diameter until both spots touched the equator. The figure shows the two groups on September 13, when the diameters of the northern and southern spots were 50,000 km. and 42,000 km., the distance between their centres being 50,000 km. The maximum intensities of their magnetic fields on that date were 3,400 Γ and 2,700 Γ. The different polarities are indicated by the letters R and V (referring to the red and violet Zeeman components of spectral lines, whose greater or less intensity indicates the sign of the magnetic field).

Fig. 17 indicates the law of sunspot polarity; the curves represent the approximate variation in mean latitude and the corresponding magnetic polarities of spot groups observed at Mount Wilson from June 1908 to January 1925. The leading (western) spot is shown on the right as it appears to an observer on the northern hemisphere of the

earth, and the opposite polarities are indicated by the letters R and V. The letters N for R-spots and S for V-spots denote 'north-seeking' and 'south-seeking' magnetism of the spot; S is therefore of the same
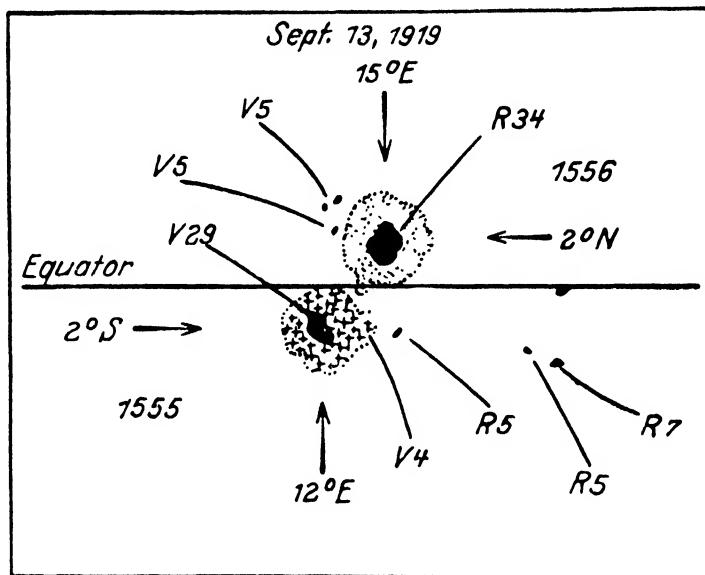
FIG. 16. Spots of opposite polarity near the sun's equator, September 13, 1919, Mt. Wilson. (After G. E. Hale and S. B. Nicholson)
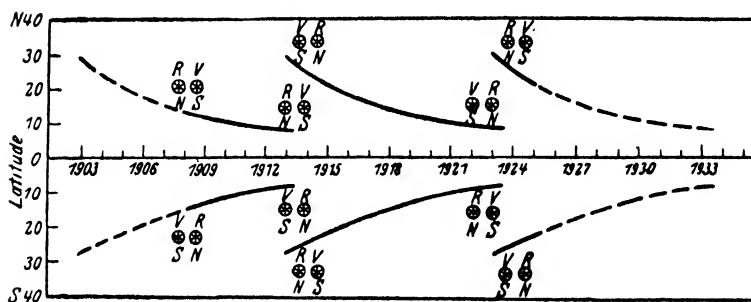
FIG. 17. The law of sunspot polarity. The curves represent the approximate variation in mean latitude and the corresponding magnetic polarities of spot groups observed at Mt. Wilson from June 1908 to January 1925. The preceding spot is shown on the right. (After G. E. Hale and S. B. Nicholson)

polarity as the magnetic pole of the earth's northern hemisphere. Fig. 18 indicates the state of affairs for about two years at the sunspot-minima of 1912 and 1923, when, in different latitudes, spot-pairs of the new and old cycles, of opposite polarities, existed.

Regions of intense magnetic polarity have sometimes been found unaccompanied by a visible spot. Hale has called these *invisible sunspots*.
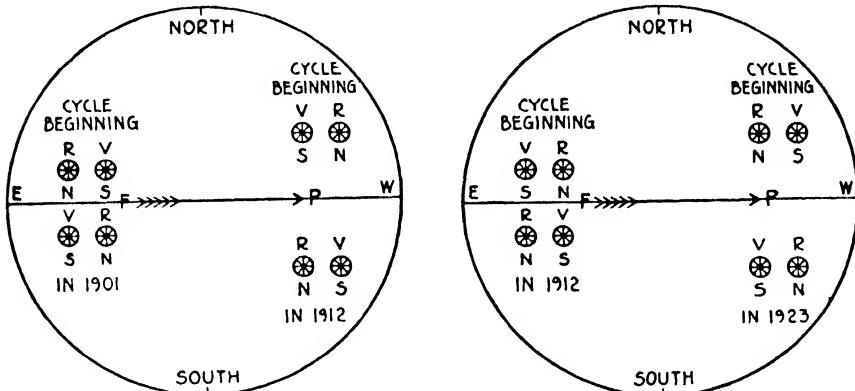
FIG. 18. Diagrams illustrating the simultaneous presence of sunspots (or sunspot zones), belonging to the new and the old cycles, for about two years near the times of sunspotminimum. In the two zones the spot groups are of opposite magnetic polarities. The left-hand diagram refers to the sunspot-minimum of 1912, and the right-hand diagram to the minimum of 1923. F = following, P = preceding (referring to the direction of motion of the spots due to the sun's rotation). (After G. E. Hale and S. B. Nicholson)

5.9. The distribution and periodicity of prominences.

The frequency of appearance of prominences undergoes a cycle closely corresponding with that shown by sunspots. The two phenomena have maxima and minima at nearly the same epochs, and, in both, the increase to maximum tends to be more rapid than the subsequent decrease; there are, however, some distinct differences between the two curves.

Ricco found that in both hemispheres there is a zone of maximum prominence frequency between 20° and 40° heliographic latitude, i.e. in the sunspot zones; these zones are active throughout the cycle, save near sunspot-minimum; the metallic eruptive prominences are nearly confined to these zones. There is another pair of zones of maximum frequency between 40° and 60° latitude; these are active during the decline from sunspot-maximum, but the latitude of activity begins to move polewards just before sunspot-minimum, and attains about 80° by sunspot-maximum. The prominences in these higher zones are usually of the quiescent type. These changes are illustrated by Fig. 19, due to W. J. S. Lockyer [10], based on Evershed's observations (1890–1920).

5.10. Periodic changes in the corona.

Lockyer's diagram (Fig. 19) also illustrates the relation between the prominences and the corona; he classified the coronal forms into three main types, polar, intermediate, and equatorial, illustrated in Fig. 20, and concluded that the
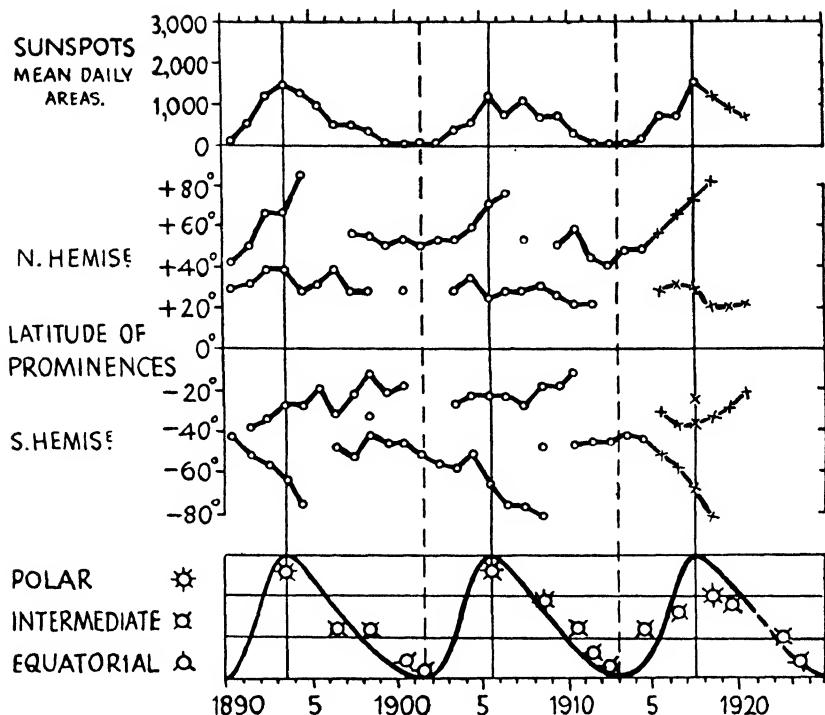
FIG. 19. Comparison between the latitudes of the prominence zones and the forms of the solar corona, 1890–1924. (After W. J. S. Lockyer.) The forms are represented by schematic diagrams corresponding to those of Fig. 20
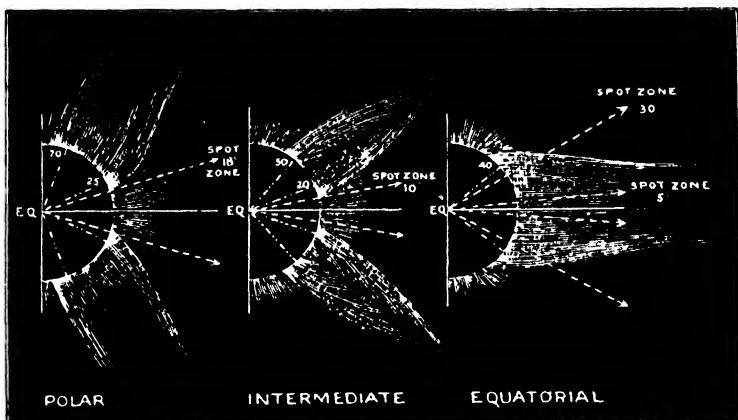
FIG. 20. Types of the solar corona (polar, intermediate, and equatorial); the centres of action of prominences are also shown, by the short radial lines. (After W. J. S. Lockyer)

various forms are clearly connected with the zones in which the centres of action of the prominences occur. The *polar* coronal form appears when the prominences are most frequent near the sun's poles; the *equatorial* type occurs when there is only one active zone, at about 45° latitude, in each hemisphere; and the *intermediate* type occurs when there are two active zones in each hemisphere, neither being near the poles. The special arched form of some streamers is attributed to the action of two prominence zones situated near their bases. Sunspot activity apparently has no direct connexion with the production of coronal streamers.

5.11. The distribution and frequency of faculae. The number of faculae (particulars of which are published in the *Greenwich Photoheliographic Results*) undergoes a variation similar to that of sunspots; the conformity in their variations hold good not merely for the spots and faculae as a whole, but also for those in each (northern and southern) hemisphere separately (cf. Fig. 21), though the progression of sunspot activity differs as to the dates and amounts of the chief developments in the two hemispheres.

E. W. Maunder [7 b] found a slight systematic predominance (about 3 per cent.) in the area of faculae on the eastern as compared with the western limb of the sun. The predominance was least during the rise to maximum activity, and greatest during the succeeding decline.

The following quotation [7 c] concisely summarizes the Greenwich observations of faculae, 1874–1917:

'The centres of the chief zones of the faculae have a well-defined progression with the solar cycle of about 11 years . . . , in considerable accordance with the latitude progression of the sunspots. . . . As compared with the spot-zones, however, the corresponding faculae-zones are on the average about 15° broader. The extension is mainly polewards, and . . . made up entirely of small areas of faculae, the largest areas being always confined to the region of the spots.

' In character, the faculae unassociated with sunspots are small faint areas lasting at the most for two months. This characteristic offers a striking contrast to the faculae connected with sunspots, which develop in a few days into bright and compact masses often covering a great area. These after about two months become faint and more scattered, but they can nearly always be discriminated from those which have at no time been seen with sunspots. With large spot-disturbances, the faculae can be recognized frequently for several months.

'The mean percentage area of faculae unconnected with spots as compared

with the total area of all faculae is about 10 per cent. This figure increases to about 30 per cent. during the minimum years and falls to 5 per cent. when the Sun is active.

'An area of very bright compact faculae indicates that either a spot-group
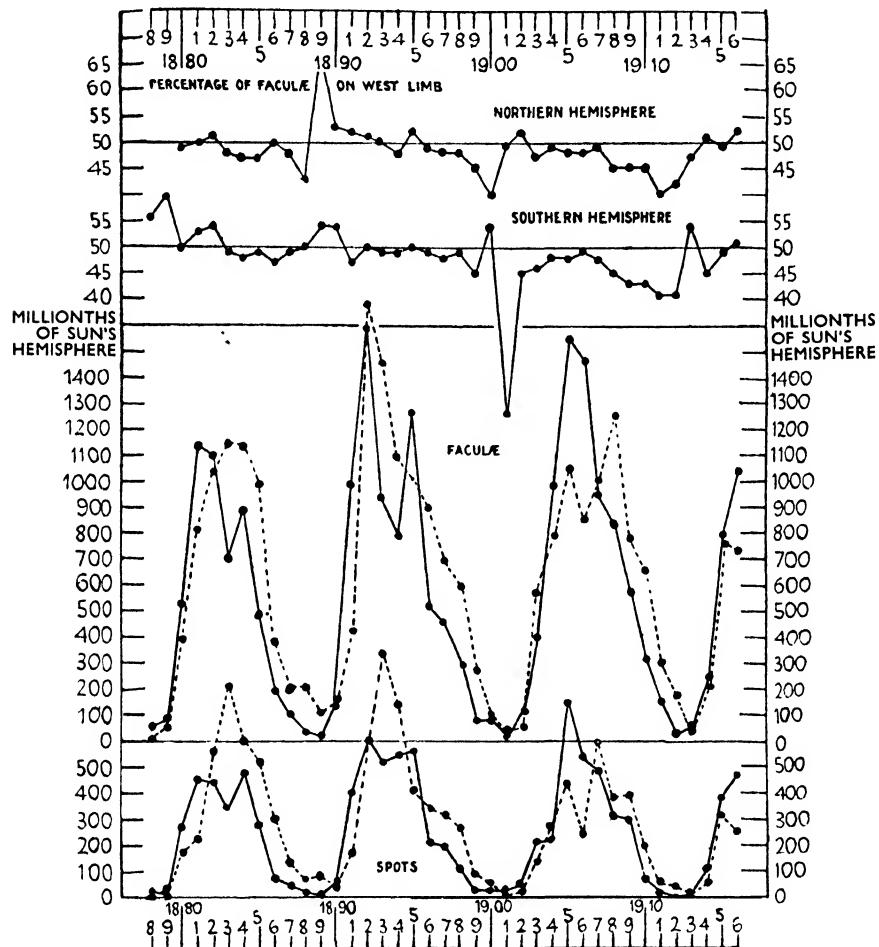
Fig. 21. The percentage of faculae seen on the west limb of the sun, and the mean daily areas of sunspots and faculae, 1866–1915, corrected for foreshortening, and expressed in millionths of the sun's visible hemisphere; the continuous lines refer to northern and the broken lines to southern spots and faculae. (After E. W. Maunder)

has been very recently connected with it or that one will appear within a few hours.

'The duration of faculae connected with a spot-group is on the average at least three times the length of the accompanying spots, but the proportion is a very variable one. Generally speaking, the faculae of large

spot-groups as compared with those of smaller groups do not last proportionally as long.

'The faculae frequently act as a connecting link between successive spot-disturbances in the same region. Near the maximum of the solar cycle, some of these centres have been traced without intermission for over six months and a few for nearly a year.

'Faculae frequently appear in streaks roughly at right angles to the direction of the Sun's rotation, and have a strong tendency to spread from a spot-disturbance for several degrees in latitude. This feature contrasts with the spots themselves which invariably stream out in longitude, whilst there is little trend in the direction of latitude.

'The faculae change form rapidly, and individual features can seldom be recognized at successive appearances at the Sun's limbs. The groups of faculae considered as entities are, however, more stable than groups of spots. . . . The regularity of statistics of sunspots is largely disturbed by the fact that we have an imperfect record of the spots owing to the Sun's rotation. The longer life of the faculae tends to smooth this.

'The investigation of the faculae has brought out more strongly than was anticipated the very close connexion between them and the sunspots. There are no spots without faculae, and no extensive areas of faculae without spots.

'There is a zone of polar faculae in both hemispheres . . . about latitude 70°, . . . small, short-lived, detached flecks. Their appearance seems to be somewhat erratic and shows no pronounced relationship with the solar cycle . . .; they are not associated with the polar prominences.'

5.12. Solar physics. The physics of the sun's atmosphere [1] is beset by many unsolved problems. The outermost layer of the atmosphere (apart from the corona) is called the *chromosphere*; this is a thick but highly rarefied layer in which the density increases downwards much more slowly than in the layer below, which is called the *reversing layer*. This merges into the *photosphere* at a level where the gas is ceasing to be transparent. The gas is largely ionized throughout the whole atmosphere.

The gas in and below the reversing layer is supposed to be supported mainly by ordinary gas pressure, but the mode of support of the chromosphere is still uncertain. Milne [15] suggested that the ionized calcium in the chromosphere is supported by selective radiation pressure, and it can be shown that for these ions this pressure is comparable with solar gravity. Another form of this theory of support by selective radiation pressure was proposed by Chandrasekhar, but according to McCrea [ref. 1 b, p. 99] it may 'easily be demonstrated that the total upward momentum available from selective radiation pressure is insufficient to support the mixture of atoms actually found in

the solar chromosphere, against solar gravity '. Dyson and Woolley [1b] consider that turbulence, as suggested by McCrea, represents the most satisfactory solution of the problem of support of the chromosphere.

It has been suggested that the rapid changes in the prominences above the chromosphere may be regarded as due to variations in the support afforded by selective radiation pressure, similar except in degree to that which supports the chromosphere. Local bright areas on the sun exert excess radiation pressure, and tend to repel matter capable of absorbing their radiations; dark areas, on the contrary, exert less than the normal radiation pressure, and this deficiency is equivalent to a relative attraction of absorbing matter towards such areas; Pike has shown how the arched forms of prominences may possibly be explicable on these lines.

Milne [14] has pointed out that chromospheric matter which happens to be moving upwards will be continuously accelerated; the wavelength of the radiation which it absorbs will be shifted towards the violet, according to the Doppler principle, by reason of the upward motion of the atoms, and if the upward-moving atoms are not numerous enough to alter the normal absorption band appreciably, they will be subject to increasing radiation pressure, and acceleration, until they absorb light in the strong continuous spectrum on the violet side of the absorption line. Their increasing distance from the sun will diminish the acceleration, owing to the operation of the inverse square law of radiation pressure, and the atoms will acquire a limiting velocity of the order 1,600 km./sec. The velocity of escape is 615 km./sec.

The nature and origin of sunspots is still far from clear. They are the seat of vortical motion, and are of much lower temperature (about 3,000°) than the surrounding photosphere; Bjerknes [13] has given an interesting theory which attempts a qualitative explanation of their remarkable collective properties.

5.13. The sun's general magnetic field. Besides the intense local magnetic fields existing in and near sunspots, the sun seems to possess a general magnetic field, discovered by Hale [4, 19–29]. Its magnetic axis very nearly coincides with its axis of rotation, the estimated inclination being only 4°; the magnetic axis is supposed to rotate relative to the sidereal system once in about 32 days. The positive direction of the magnetic axis is related to the direction of solar rotation in the same way as for the earth. But the sun's magnetic field differs greatly from that of the earth in that its intensity, which at the equator is about 40 Γ at the lowest level at which (by the Zeeman effect) it can

be measured, decreases upwards rapidly, to 10 Γ, the lowest measurable value; the range of height in which this decrease occurs is not more than 300 km. There is no reason tò suppose that the decrease ceases beyond 10 Γ. Deslandres, from the study of spiral motions of matter in prominences, concludes that above the chromosphere the magnetic intensity is of the order 10^{-7} Γ.

The rapid radial decrease of the magnetic intensity in the reversing layer of the sun (i.e. the layer, just above the photosphere, which is supported largely by gas pressure) indicates that the lines of magnetic force must be nearly horizontal in the atmosphere, in the south-to-north direction. But all these statements about the sun's field are still uncertain.

The form of the coronal plumes suggests that the sun is surrounded by a magnetic field extending to at least 100,000 km. from the surface; but at present the intensity of this field is quite unknown.

5.14. Sources of solar data. It is of interest, in connexion with the investigation of relationships between solar and terrestrial phenomena, to indicate some of the main data referring to the state of the sun on each day. The relative daily sunspot numbers, and daily sunspot areas, have already been referred to (§ 5); the former are published in special bulletins (*Astronomische Mitteilungen*) issued by the Zürich Observatory, and also in Terrestrial Magnetism, the *Meteorologische Zeitschrift*, and the *Quarterly Journal of the Royal Meteorological Society*. Diagrams showing the forms of prominences observed spectroscopically (in Hα light) at the limb of the sun are published at intervals by the Arcetri Observatory, under the auspices of the International Astronomical Union (the last issue, in 1939, referred to the years 1933 and 1934); this publication (entitled *Immagini spettroscopiche del bordo solare osservate a Catania, Madrid e Zurigo*) also tabulates the mean daily areas of prominences for every 5° of latitude, in a unit equal to 1° of the solar circumference in length, by 1 second of arc of the celestial sphere in height. The Kodaikanal Observatory publishes every six months the mean daily areas (in square minutes of arc), and the numbers, of prominences recorded photographically at various observatories. The Paris (Meudon) Observatory, also under the auspices of the International Astronomical Union, publishes synoptic charts of the solar chromosphere, with a catalogue of high-level filaments.

The quarterly *Bulletin for Character Figures of Solar Phenomena*, published, for the International Astronomical Union, by the Zürich Observatory since 1928, and extended backwards to 1917, contains

character-figures on the scale of 0 (absence or rarity), 1, 2, 3, 4, 5 (extreme abundance and intensity) for the *whole sun disk* and for a *central area* (since 1929 this has been the central part of the disk, having a radius half that of the whole disk), for Calcium-Flocculi and Bright and Dark Hydrogen-Flocculi; also Relative Sunspot Numbers for the whole disk, the central area, and the four quadrants divided from the north pole, and a list of brilliant chromospheric eruptions and active regions, collected by L. d'Azambuja. The central zone is selected as favourably situated, on the day in question, for influencing the earth by means of radial, or nearly radial, emissions. However, at the Stockholm Assembly (1938) of the International Astronomical Union, it was decided to discontinue the data for the central area (and for the four quadrants), after December 1938; the title also was changed to *Quarterly Bulletin on Solar Activity*.

The *Astronomische Mitteilungen*, founded by R. Wolf, continued by A. Wolfer, and now edited by W. Brunner, give annual statistics of sunspots, faculae, and prominences, and complete heliographic charts of spots and faculae.

The *Greenwich Photoheliographic Results* give areas of sunspots and faculae and other tables in annual volumes.

Cosmic-data 'Ursigrams' (*Union-Radio-Scientifique Internationale*) are collected by Science Service, Washington, D.C., and broadcast daily, describing, in code form, magnetic and solar phenomena and ionospheric data. Codes and transmission schedules have been given by Watson Davis [G 93]; the data are published quarterly in *Terrestrial Magnetism*.

5.3. Additional notes. Newton [11.31 a] gives the following as the percentages of bright solar eruptions of the classes 1, 2, 3 respectively: 77, 19, and 4. He also remarks that though these bright eruptions suggest the 'disturbed regions . . . of brightness fluctuations' from which, according to Milne's theory, high-speed calcium atoms are most likely to be ejected from the sun, yet there is at present little observational evidence for this, because the outward velocities shown by the absorption markings rarely equal or exceed the velocity of escape from the sun. Barely half a dozen such cases have ever been recorded.

Newton has compiled a list of 91 bright eruptions of intensity 3, from spectrohelioscopic records, and by a search of published spectroheliographic data since 1892; of these 91, there were 29 so intense as to be classed as 3+. These are further discussed on p. 395.

VI

A GENERAL REVIEW OF THE TRANSIENT MAGNETIC VARIATIONS

6.1. The transient magnetic variations, S, L, and D. The continuous magnetic records of any observatory show that on some days all the three elements undergo smooth and regular variations, while on other days their changes are more or less irregular. Days of the first kind are said to be magnetically *quiet* or *calm* (at that station); days of the second kind are called magnetically *active* or *disturbed* days. When extreme, a magnetic *disturbance* is called a *magnetic storm*.

The degree of magnetic *activity* or disturbance varies from day to day over a wide range, and few days seem wholly free from disturbance. But except in periods of extreme activity it is evident from the magnetograms that the disturbance is superposed on a *regular daily variation* (or *diurnal* variation, though in this book 'diurnal' will usually be used only in the special sense of a *harmonic* variation with a period of one day). This may be called the *solar daily variation*, to distinguish it from another regular daily variation which the magnetic field is found to undergo, depending on lunar time; this is called the *lunar daily variation*. For brevity it is convenient to refer to these two daily variations and to their respective fields by the letters S (solar) and L (lunar). The solar daily variation is seen in its pure form on quiet days, and it is convenient to denote it by S_q. The letter D may likewise be used to refer to magnetic disturbance and its field. These three variations, S, L, and D, together with the secular variation, comprise the chief changes in the earth's magnetic field. Since S, L, and D, unlike the secular variation, produce no really long-enduring change in the field, they will be called the *transient* magnetic variations.

One important distinction between S and D on the one hand, and L on the other, is that the former can be recognized at sight from the magnetograms, whereas in general the lunar daily variation L cannot (but cf. 8.16), because its range is small; it can be determined only by averaging over many days. But its small magnitude does not detract from its great theoretical interest.

Each magnetic element is affected in a characteristic way by each of the three types of variation S, L, and D. Hence a single observatory provides material for nine separate investigations—on the nature of S, L, and D in each of the three elements.

But to gain a comprehensive understanding of the transient magnetic variations it is necessary to relate and analyse (and also synthesize) the results of such studies from many observatories widely distributed in latitude and longitude. Moreover, this work should be done not only for S, L, and D as determined from the mean of one or more years, but also for each season separately, because both the type and the range of these variations are found to alter in the course of the year. (In order to examine the seasonal changes in these variations, their monthly means, or their averages for the three groups of four months each (5.1), June solstice, December solstice, and equinoxes, should be considered.) Further, the range of the variations, and in the case of D its frequency of occurrence, vary from year to year, requiring an extension of the investigation to different years. The study of the transient magnetic variations is thus an extensive, complex, and laborious task. It is also not easy, when the chief facts have been found, to describe them clearly and concisely. Only gradually, over a long period of time, have suitable methods of investigating and representing them been found—methods likely also to be of great value in allied branches of geophysics that have been more recently developed.

In this chapter a general review of the transient magnetic variations will be given, because they are to some extent inter-connected, and it is difficult to describe any of them in detail without implying some knowledge of the others.

6.2. The magnetic classification of days. The distinction between quiet days, when the magnetograms are smooth and simple in form, and disturbed days, when they show irregular fluctuations (or, if simple, are abnormal), has already been illustrated in 2.9 : it is naturally not precise. No practicable method of arriving at a numerical measure of magnetic activity founded on a clear physical basis has yet been formulated. Several rough measures, however, have proved convenient and successful for certain purposes. The simplest is the daily range—the difference between the highest and lowest value recorded—in a single element, usually D or H: for days falling within a short period, such as a month, this affords a fairly good classification of days according to their relative degree of disturbance, the days of greater disturbance generally having the larger range. But for several reasons the simple ranges cannot be regarded as numerical measures of the degree of disturbance; they do not become zero on even the quietest days, and they show a more or less pronounced annual variation. The range also on some days may depend on a single brief variation in the

field, and on the time of day when it occurs: it alone will not distinguish between such days and others throughout which there is continual disturbance of similar intensity. This can be partly remedied by taking account, in judging the relative degree of disturbance, of the general appearance of the magnetograms (preferably all three), as well as their ranges. In this way, at any observatory, a rough classification of each day, or other convenient interval, can be made, according to the magnetic activity then prevailing.

By international agreement such a classification has been made since 1906, for each Greenwich day (2.13), at a number of co-operating observatories (now about forty). Recently this classification has been extended backwards to years before 1906. The classification at each individual station is crude, being only into one of three classes, whereas actually there is a continuous gradation of disturbance over a wide range: the three classes are designated by figures, 0 for a quiet day, 1 for an average day, 2 for an unusually disturbed day. For each day the mean of the figures supplied by all the co-operating observatories is taken, to one decimal place, and the result is called the *international magnetic character-figure*, C, for the day [4]: these figures naturally afford a more delicate classification of days, since there are 21 classes, corresponding to the numbers 0.0, 0.1,..., 2.0.

The daily character-figures are valuable because they classify Greenwich days approximately according to the magnetic activity averaged over a large part of the world, although they do not represent a numerical measure of any clearly defined physical entity. On account of the greater density of observatories over Europe, the state of disturbance there has undue weight in determining the character-figure for the day. Again, the figures from individual observatories are intentionally (cf. [2]) not assigned on any clearly formulated plan; the primary object of the scheme when instituted was to distinguish adequately between the days of a single month, so that a proper choice of the five quietest days per month (now called the *international quiet days*) might be made, to serve in obtaining a mean representation of the changes on quiet days (more recently the daily character-figures have also been made the basis of choice of five *international disturbed days* per month). Hence the figures assigned at any observatory have reference to the relative amount of disturbance on different days at that station (as some stations experience more intense variations than others, the average amount of disturbance at one station might be an extreme amount at another). The practice in assigning the figures 0, 1, 2 may also vary

from one month or year to another at the same observatory, the standards of quietness for the figure 0, or of disturbance for the figure 2, being modified according to the general prevalence or absence of disturbance. Attempts are being made to devise a simple practicable plan of measuring the magnetic activity which will remove or reduce some of the disadvantages of the present scheme: the latter, however, has proved very useful [2, 3]. (See also p. 213.)

6.3. World-wide simultaneity of magnetic activity. Table 1 gives the daily character-figures, and an indication of how many 0's, 1's, and 2's were assigned by different observatories to each day, for a single month, November 1936, including both quiet and disturbed days. It shows a considerable degree of accordance between the figures assigned for the same Greenwich days at different observatories, especially for the days of high or low character-figure (for instance, the 29th or the 23rd). It suggests that the incidence of disturbance is world-wide, though the distribution of relative intensity is not uniform over the earth. The demonstration is limited, however, to the middle

TABLE 1. *International Daily Magnetic Character-figures for November 1936*

Day	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
No. of 0's .	38	14	0	10	13	19	14	15	7	8	0	19	50	38	2
„ 1's .	14	37	23	36	36	30	30	33	37	42	21	32	2	13	35
„ 2's .	0	1	29	6	3	3	8	4	8	2	31	1	0	1	15
Char.-fig. .	0·3	0·8	1·6	0·9	0·8	0·7	0·9	0·8	1·0	0·9	1·6	0·7	0 0	0·3	1·2

Day	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
No. of 0's .	3	11	5	9	26	38	50	52	52	49	24	47	26	0	29
„ 1's .	44	37	41	38	25	14	2	0	0	3	25	5	18	1	23
„ 2's .	5	4	6	5	1	0	0	0	0	0	3	0	8	51	0
Char.-fig. .	1·0	0·9	1·0	0·9	0·5	0·3	0·0	0·0	0·0	0·1	0·6	0·1	0·7	2·0	0·4

Five quiet days: 13, 22, 23, 24, 25.   Five disturbed days: 3, 11, 15, 16, 29.

belt of the earth, since almost all the contributing observatories lie between 60° N. and 32° S., the majority, indeed, between 53° N. and 20° N. The evidence afforded by the table is therefore valuably supplemented by an investigation due to Chree [5], in which the international character-figures were compared with character-figures for the Antarctic station of the Second British National Scott Expedition of 1911–12.

These figures 0, 1, 2 were assigned independently by Chree by inspection of the Antarctic magnetograms, with sole reference to the more

or less oscillatory nature of the curves: the standards of quietness and disturbance were chosen so that the relative proportion of days of characters 0, 1, and 2 should be about the same as in the international tables: otherwise, owing to the high average disturbance at this, as at other, polar stations, too few 0's and too many 2's would have been assigned. The following table gives some particulars of the distribution of *Antarctic* character-figures on the five international quiet days per month (110 in all) during the available period of the Antarctic observations—about 22 months—and likewise for the five international disturbed days per month (107 in all).

	No. of Antarctic character-figures equal to			Mean Antarctic figures
	0	*1*	*2*	
107 international disturbed days .	0	22	85	1·8
110 international quiet days .	57	45	4	0·5

The table shows clearly 'that days of large disturbance in temperate latitudes are always, or almost always, days of large disturbance in the Antarctic: while days that are conspicuously quiet in temperate latitudes are usually quiet, and very rarely much disturbed, in the Antarctic'. 'Three of the four international quiet days which received Antarctic figures 2 occurred in December 1911, a month to which only one 0 and sixteen 2's were assigned: all three were on the borderline between 1's and 2's, being essentially quiet, save during a few hours near midnight.'

It may therefore be concluded that a day with a high international figure is highly disturbed almost all over the world, and one with a low character-figure is almost everywhere relatively quiet. The records reproduced on pp. 50–1 are typical examples.

6.4. The geographical distribution of the average intensity of magnetic activity. Though every period of magnetic disturbance has an individual or accidental character of its own, the distribution of intensity of disturbance over the earth has some well-marked average properties. The distribution is in general related to two variables of position, the latitude l (or, in high latitudes, the magnetic latitude), and the longitude λ' relative to the sun, that is, the local time. Regions in high (magnetic) latitudes are generally much more disturbed than those nearer the equator: the increase of disturbance with latitude is very rapid near magnetic latitude $\pm 67°$, as the zones of maximum auroral frequency (Ch. XIV) are approached. Within these zones the

distribution of disturbance-intensity is less thoroughly known, but the intensity appears to be high.

The variation of the disturbance-intensity with respect to the local-time longitude λ' is less pronounced, and less is known about it; the value of λ' at which the disturbance-intensity attains its maximum is different for different latitudes; Stagg has shown that there is a notable change in the value of λ' for the maximum disturbance-intensity on crossing the auroral zone (11.19).

6.5. The time-pattern of occurrence of magnetic disturbance. The incidence and duration of magnetic disturbance appear largely fortuitous, and cannot be predicted with any certainty, though there is a tendency for disturbance to recur after about 27 days (Ch. XII); but certain statistical regularities have been found in them, which indicate a close connexion between magnetic disturbance and the sun (Chs. XI, XII). Periods of magnetic quiet may last for several days, but rarely for more than a month. They may end by the gradual growth of disturbance, or by the quite sudden outbreak of an intense disturbance or storm. The most intense storms nearly always *commence* suddenly—at the same time, to within a minute or so, all over the earth. Strong disturbance rarely *ends* abruptly, however; it gradually subsides, and it is rare for a highly disturbed day to be followed immediately by a quiet day. On the other hand, disturbance does not remain at a high intensity for long—sometimes only for a few hours, and rarely for more than three or four consecutive days. Frequently a quiet day follows within three days of a large disturbance.

The statistical relationships between magnetic disturbance and the sun are of various kinds, some depending on the geometrical factors described in 5.1, and others upon the intrinsic physical state of the sun. The geometrical factors depend on (i) the position of the earth in its orbit, which governs the seasonal variations in the distribution of disturbance-intensity over the earth (11.9); and (ii) the orientation of the sun about its axis, which determines a remarkable tendency for disturbed conditions (or their absence), at any time, to recur after the lapse of one or more rotation periods of the sun (Ch. XII). As to the annual variation in the intensity and frequency of disturbances, with maxima at the equinoxes, it is not yet known whether this is due to the geocentric position of the sun or to the heliocentric position of the earth at those seasons. The connexion between magnetic disturbance and the intrinsic state of the sun is indicated by a marked synchronism between the changes in the average amount of magnetic

disturbance from year to year, and the changes in the corresponding annual mean sunspot numbers or areas (Ch. XI).

Magnetic disturbance also has associations with other terrestrial phenomena similarly correlated with solar changes, and notably with aurorae, earth-currents, and disturbances of the ionosphere (Ch. XV) such as are measured by radio methods.

6.6. The general character of a great magnetic disturbance. A great magnetic disturbance, especially when its beginning is clearly defined, has a certain unity, and many different storms show a broad resemblance to one another in their progress and in the distribution of their variation fields, although each has also many irregularities peculiar to itself. One marked feature is that throughout middle and low latitudes a magnetic storm usually reduces the daily mean value of H. This reduction disappears only gradually, and the normal value is often not restored until some days after the ordinary signs of disturbance—irregular fluctuations—have ceased. Thus certain disturbance effects persist even during quite quiet days following a magnetic storm; van Bemmelen termed this the *post-perturbation (Nachstörung)* or residual disturbance. It shows itself in H by a small daily rise (the 'recovery'), contributing, along with the secular variation and any irregular changes, to the non-cyclic variation (2.14 and 16.2).

6.7. The solar daily magnetic variations on quiet days. The daily inequalities of the magnetic elements at any station, derived from quiet days, and freed from the residual effect of disturbance represented by the non-cyclic variation, may be regarded as defining for that station the components of a special variation field, denoted by S_q. The whole field S_q is determined by such inequalities from all stations, and in the first instance is defined only during quiet periods. The distribution of the S_q field undergoes a considerable annual variation, but at any particular epoch in the year it is of a fairly definite simple type; this alters more or less from day to day, and also its intensity as a whole fluctuates somewhat, in an irregular manner.

The daily inequalities of X, Y, and Z for stations in the same latitude but in different longitudes are approximately the same at corresponding local times. This implies that the distribution of the S_q field is a function of the local-time longitude λ' and the co-latitude θ, and is approximately constant as viewed by an observer on the sun. Hence the solar daily variation of the whole field F at each terrestrial station on quiet days may be pictured as caused mainly by the motion of the station, as the earth rotates, through the non-uniform field S_q, stationary as

viewed from the sun, except that it is carried along with the earth in its orbital motion (see however, 7.10). The intensity of the field changes irregularly, even in the course of a single month, but on the average approximately in the same ratio everywhere.

On account of the regular seasonal change in the S_q field, it is convenient to consider the average value of S_q during each calendar month (or during the three seasons j, e, and d of 5.1). This is defined by the mean daily inequalities derived from five well-distributed quiet days in each calendar month, or from several sets of such days from the same calendar month in different years. The choice of such days is now made in connexion with the international scheme (6.2) for the magnetic characterization of Greenwich days; when quiet-day variations (from 1906 onwards) are referred to, it is to be assumed, unless otherwise stated, that they are derived from the international five quiet days. Before 1906, from 1889 onwards, five quiet days per month were chosen with the same object by the Astronomer Royal; but the magnetic characterization of days, and the choice of five quiet and five disturbed Greenwich days per month, is now being extended backwards (see Tables G and H in Vol. II). The remaining days, which are not selected as quiet or disturbed, will be referred to as *ordinary* days.

Though the monthly mean variation field thus derived from five quiet days is usually taken as representing S_q, the field of the quiet-day solar daily variation, actually it differs slightly from S_q on any individual quiet day, not only because of the possible variation of amplitude and the slight seasonal change of type of S_q from day to day, but also because on any individual day the 24-hour inequalities will include the lunar daily magnetic variation, present on all days whether quiet or not; this will be nearly eliminated from the mean of five well-distributed quiet days, so that usually any residual part of it still present in S_q will be negligible.

The intensity of the field S_q varies not only with the season, that is, with the position of the earth in its orbit, but also with the intrinsic physical state of the sun; in sunspot-maximum years it is much more intense than at sunspot-minimum. It also undergoes day-to-day variations even during quiet periods, affecting especially the intensity of the S_q field.

6.8. The S_D field. On ordinary days, or on days of minor disturbance, marked only by a few brief fluctuations, it is evident from the magnetograms at most stations that the S_q field is still present, though some disturbance is superposed. If the daily inequalities are

computed from all days of a month, or from the five disturbed days in each of several months (the same calendar months of different years), in order, in the latter case, to average out the irregular fluctuations, the daily variation S thus obtained is found to differ systematically from S_q. The difference is attributed to disturbance; it is represented by the daily inequality formed by subtracting corresponding hourly values of the S_q inequality (in each element, for each station) from those of the whole S inequality (for the same month and sunspot epoch). The inequality thus obtained is called an S_D inequality, the suffix D being added to indicate that this additional daily variation is due to disturbance. It is called the *disturbance daily variation*.

It is the manifestation, in each element at each station, of a world-wide field (part of the D field) which will be called the S_D field. Its intensity varies with the intensity of general disturbance, and is greater when the S inequalities from which the S_q inequalities are subtracted are those for the five disturbed days per month than when they are for all days of the month considered. Its type, however, is the same in either case. Like the S_q field, its distribution depends mainly on latitude and local time, that is, its form is approximately constant relative to the meridian plane containing the sun; its distribution is, however, very different from that of the S_q field.

The fact that the daily magnetic variation S 'is in fact constituted by two variations superimposed upon each other, having different laws, and bearing different proportions to each other in different parts of the globe', was recognized (and expressed in these words) as early as 1851 by Sabine [8]. It seems later to have fallen out of mind, until it was brought to light again by the studies described in 9.1–4.

6.9. The disturbance field D. The disturbance field D is defined as the difference between the whole field at any instant, and the average field (for a period of a month or more), after allowing for the S_q and L fields. Thus, in the notation of 1.6,

$$D = \Delta_i F - S_q - L,$$

where the interval i is a month or more; L is so small at most stations that it can usually be ignored in this equation, which therefore becomes

$$D = \Delta_i F - S_q.$$

Thus in each element at each station the variation of D from hour to hour is taken to be given by the succession of hourly values of the element, less the monthly mean for the element (this difference repre-

sents $\Delta_i F$), and also less the value, for the corresponding hour, of the S_q inequality for the month.

This definition of the disturbance field is not perfect, as the resulting field D is discontinuous by a small amount from month to month, because of the change of the monthly mean and of the S_q inequality; it also includes a small part due to the secular variation and the seasonal change in S_q during the month; but it is perhaps the simplest roughly satisfactory definition of the D field that can be devised. The field is usually small on quiet days, while during periods of disturbance its intensity changes rapidly as a function both of position and time.

It should be added that though in most months five fairly quiet days can be found on which in middle and low latitudes the amount of disturbance-variation is small compared with the variation S_q, this is more rarely the case in high latitudes; even relatively slight disturbance may there be more intense than the S_q field. Hence in some months it may be preferable, in deriving the field D for polar stations, to use daily inequalities derived from less than five days, choosing the quietest days in the month (or even quiet days in the same calendar month of other and quieter years—the inequalities also possibly being corrected by some factor to allow for the variation of S_q in the course of the sunspot cycle (§ 7)).

6.10. The annual variation of the magnetic field. Unlike the meteorological elements, the monthly averages of the magnetic elements show only a small annual variation. Its amplitude is of the order of 10γ; for a good determination of it, it is necessary to average the results of several years' observations. Part of it seems to be related, in the manner described in 3.11, to the quite noticeable annual variation of magnetic activity, because it also shows a double wave, with two extremes near the equinoxes and the other two near the solstices [10].

The *daily variations* of the field (both S and L) show, on the other hand, very pronounced annual variations.

6.11. A conventional subdivision of the transient variations. A schematic subdivision of F according to the time-scale appropriate to various more or less conventionally defined magnetic phenomena uses time-averages for various intervals, ascribed to the middle of the interval. For instance, the field for a given year (sometimes called the *permanent* field, despite the secular variation) is defined by the geographical distribution of the annual mean values of the magnetic elements for that year. The time-changes of the instantaneous value

of F at any station can be approximately apportioned to the various phenomena by forming the successive deviations from means over successively shorter intervals, as in the following sequence:

The instantaneous value		F
= the secular value (the average for 10 years) . . .	=	F_0
+ (the annual mean minus the secular value) . . .	+	F_1
+ (the monthly mean minus the annual mean) . . .	+	F_2
+ (the daily meàn minus the monthly mean) . . .	+	F_3
+ (the hourly mean minus the daily mean) . . .	+	F_4
+ (the instantaneous value minus the hourly mean) . .	+	F_5

We may call F_0 the intensity of the permanent field, and its changes give the true secular variations; F_1 gives (very nearly) the average effect of the 11-year sunspot cycle on the annual means (3.11); F_3 gives the annual variation, F_4 the daily (solar and lunar) variations, and F_5 the short-period disturbance effects. The average effect of disturbance is best expressed by $(F_2 + F_3) =$ (the daily mean minus the annual mean), as given by Schmidt for Potsdam and by some other observatories (see p. 63).

6.12. Relations of magnetic to meteorological phenomena. Magnetic variations are closely related to aurorae and solar phenomena, and the theory of the daily magnetic variations connects these with periodic winds in the upper atmosphere. This might be thought to render it not unlikely that noticeable geomagnetic effects would be associated also with changes in the weather conditions near the ground.

E. Leyst [7] extensively studied the differences between the daily magnetic variations on days with high and with low atmospheric pressure. J. M. Stagg [6] found an indication that the daily variation of pressure at Aberdeen, Scotland, was not the same on magnetically disturbed as on quiet days. However, none of the suggested relations has so far survived such statistical tests for reality as will be described in Chapter XVI.

6.13. Examples of magnetic records. The consecutive hourly means published in observatory year-books provide a fairly good abstract of the magnetic records, and give a good idea of the main features of the magnetic variations if they are plotted as in Figs. 1–3. For this purpose an appropriate relative scale has first to be chosen. In ordinary photographic records the scale-value of the ordinates is, for instance, $2\gamma/$mm., and the time-rate is 15 mm. per hour. The corresponding *relative scale-value*, which determines the appearance of the curve and does not change with ordinary photographic reproduction, is $30\gamma/$hour,

meaning that the same distance which, measured along the axis of abscissae, represents 1 hour, corresponds to an ordinate which represents 30γ. This relative scale-value, however, if used in a diagram representing the magnetic variations over many days, would give an unimpressive curve; in Figs. 1–3, 2γ/hour was used and found satisfactory.

Figs. 1 and 2 refer to Niemegk observatory (near Potsdam) [12] for the first half of 1938, a sunspot-maximum year. The hourly means are plotted with respect to a zero-level chosen to correspond with the average level for the night values during quiet conditions, at a time sufficiently long after a magnetic storm; this zero-level changes with the secular variation. The days are arranged in groups (called solar 'rotations') of 27 days (see Ch. XII), with 1 day added at the beginning and 2 days at the end, so that in each compartment (upper, middle or lower) corresponding to one rotation, the curves for X, Y and Z, which extend across the double page from left (a) to right (b), represent the magnetic variations throughout 30 days. In all, nearly 12,000 hourly means are plotted in Figs. 1, 2. The divisions between the Greenwich days are indicated by lines corresponding to Greenwich midnight; the epochs of *local* noon are indicated by vertical broken lines. Horizontal lines are drawn 20γ apart. The vertical distances between the zero-lines had to be varied in order to illustrate clearly the variations during the most intense storms; for January, the curves partly intersect, while for April and May, in the parts where the departures were greatest, the curves are turned through $90°$ into the horizontal.

A close inspection of the diagrams will show many characteristic features; the diagrams should be inspected again after the reader has read Chapters VII–XII. Only a few points will be mentioned here.

The quiet and disturbed times can be clearly distinguished. The great storms are most striking: 1938 January 16/17, 22, and 25/26; April 16; May 11/12. During the nights overlapped by these storms they were seen to be accompanied by aurorae. All the five storms occur in the 'left' half of the rotations. The change of the quiet-day solar daily variation S_q from winter to summer is well marked, as well as the contrast between S_q and S_D, especially in Z. The depression of the north component X after storms, and the slow recovery, is typical.

Fig. 3 shows, in the same way, simultaneous records for 1927 October 5 to November 3 from three observatories, one in the northern hemisphere (Seddin, near Potsdam), one near the magnetic equator (Huancayo,
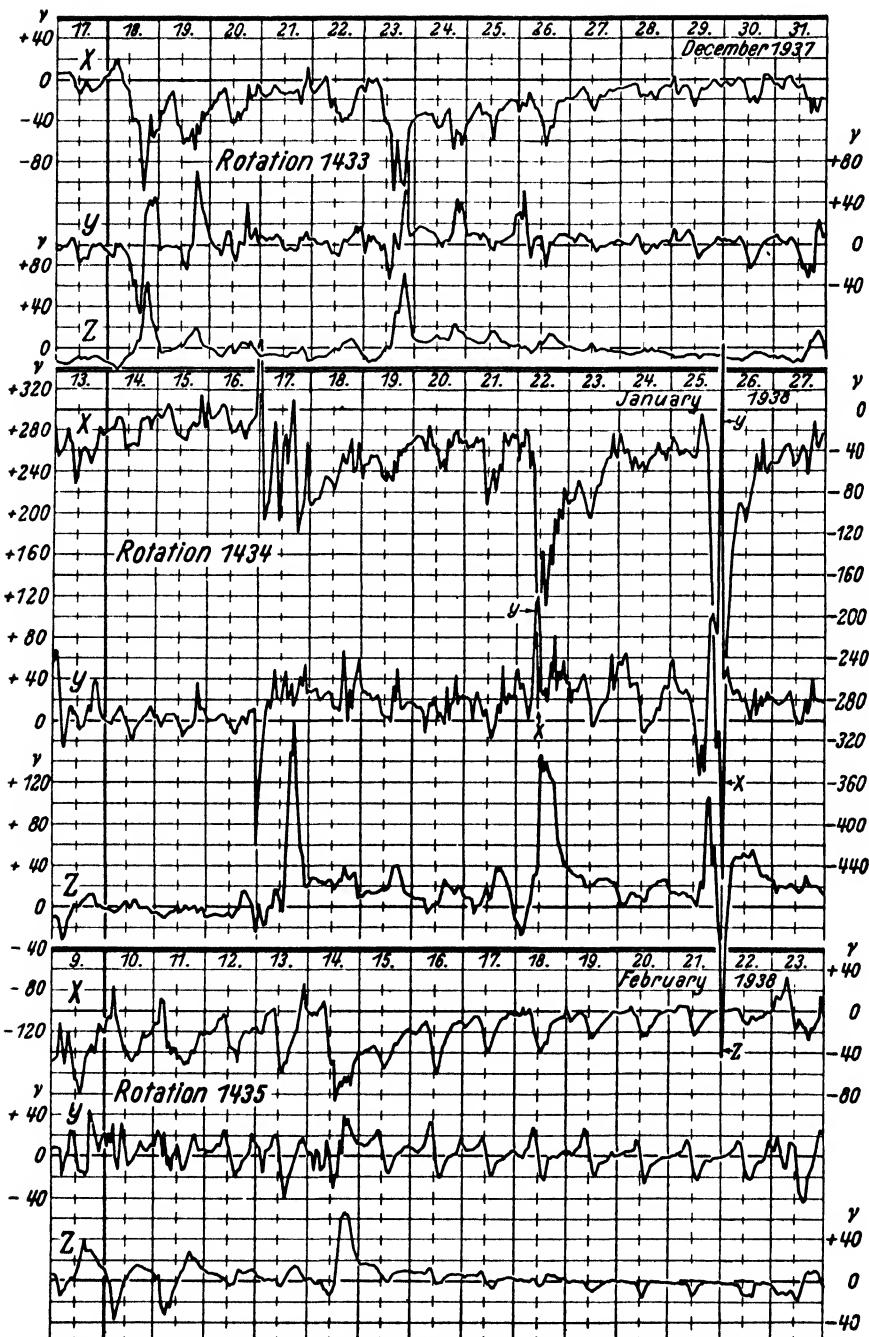
FIG. 1 A. Consecutive hourly means of the three force components at Niemegk
Observatory, near Potsdam (continued in Fig. 1 B; see description, p. 205)
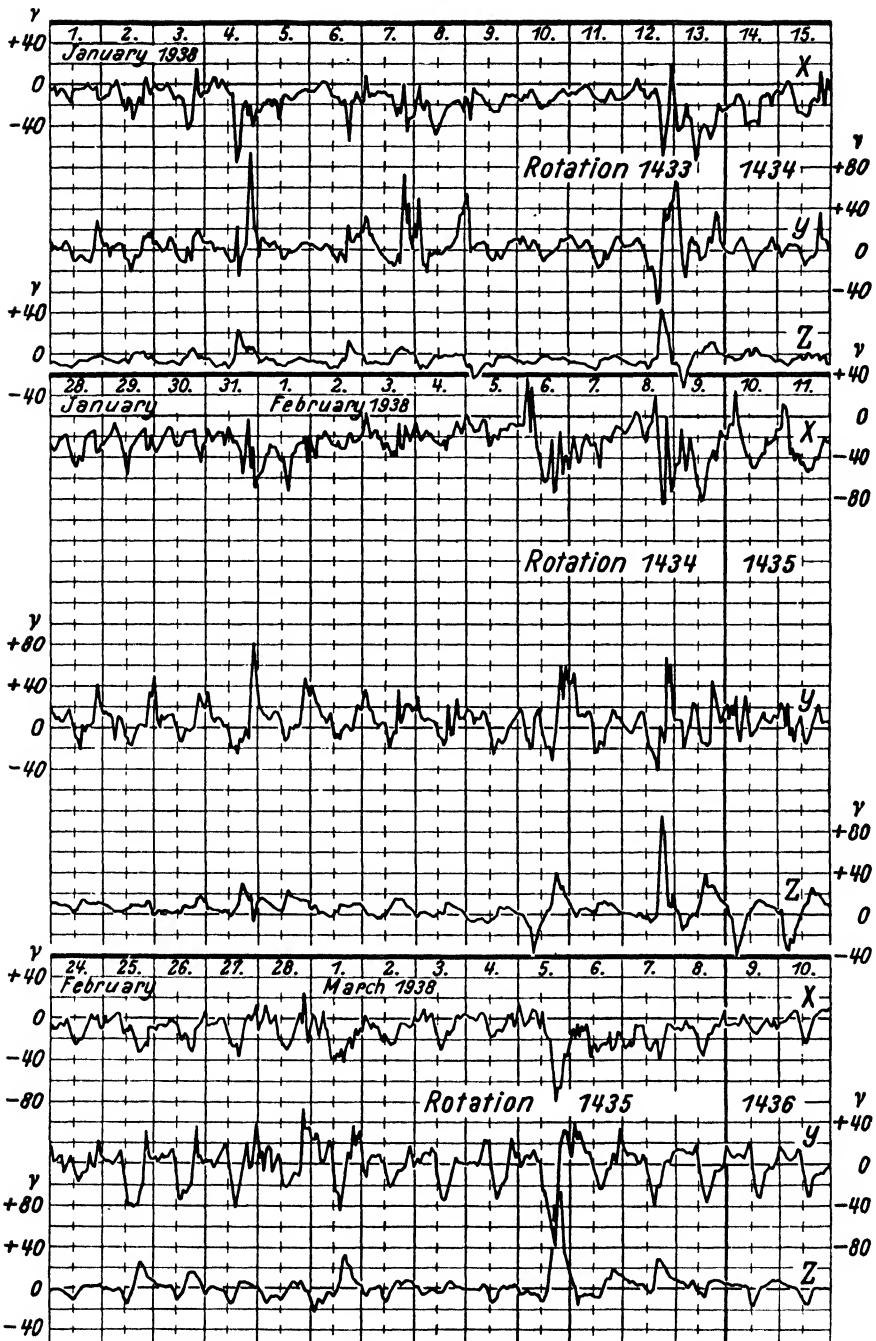
FIG. 1 B. Consecutive hourly means of the three force components at Niemegk Observatory, near Potsdam (continued from Fig. 1 A; see description, p. 205)
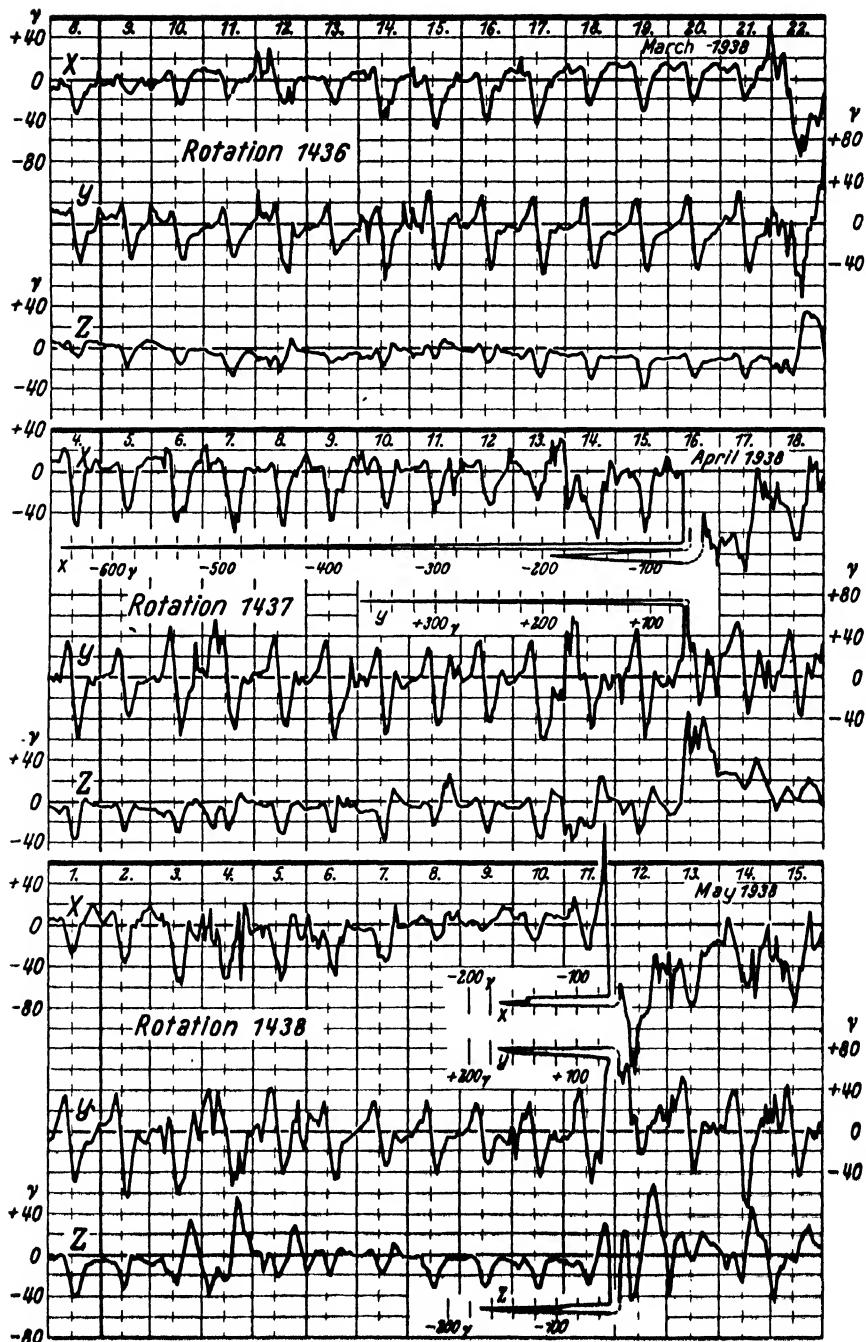
Fig. 2 A. Consecutive hourly means of the three force components at Niemegk
Observatory, near Potsdam (continued in Fig. 2 B; see description, p. 205)
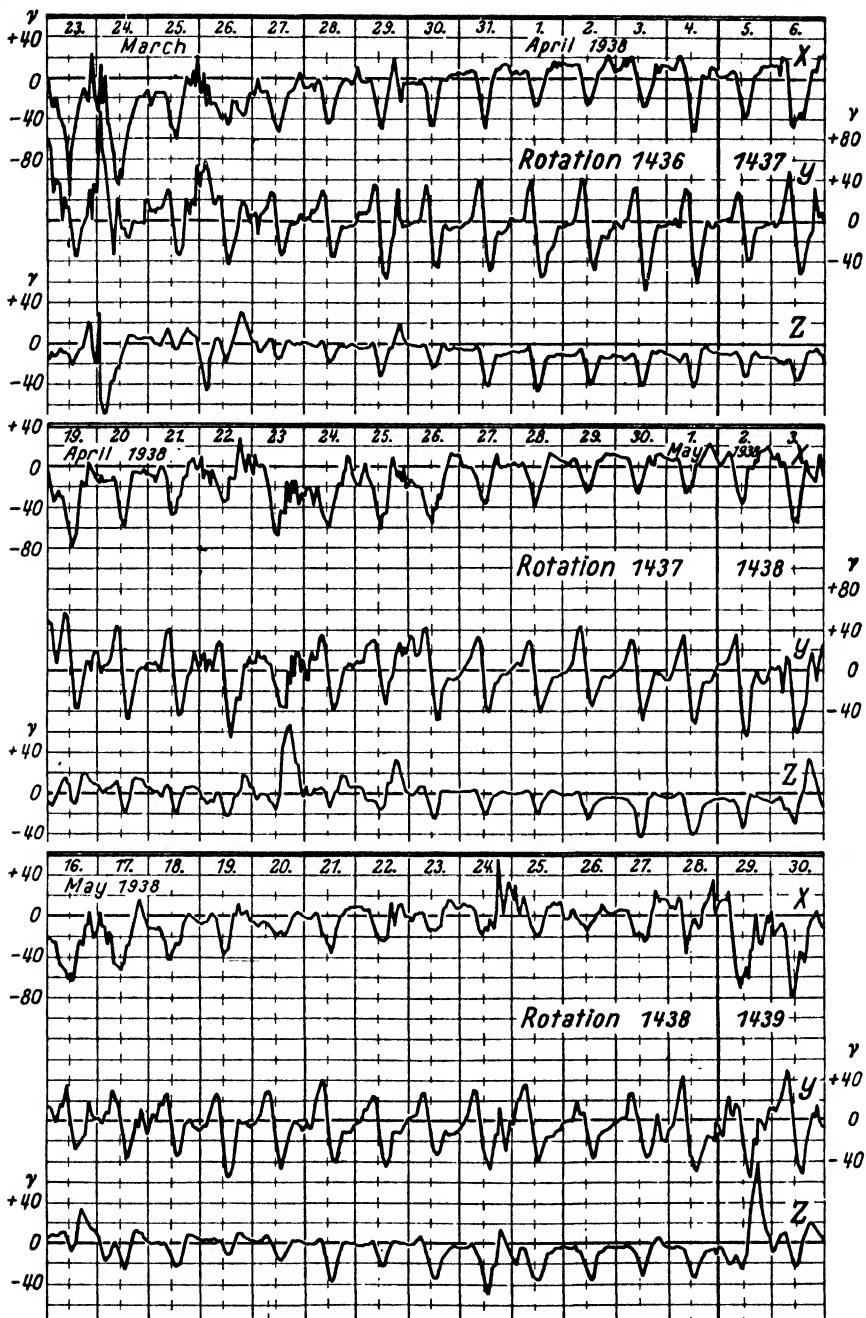
FIG. 2 B. Consecutive hourly means of the three force components at Niemegk
    Observatory, near Potsdam (continued from Fig. 2 A; see description, p. 205)
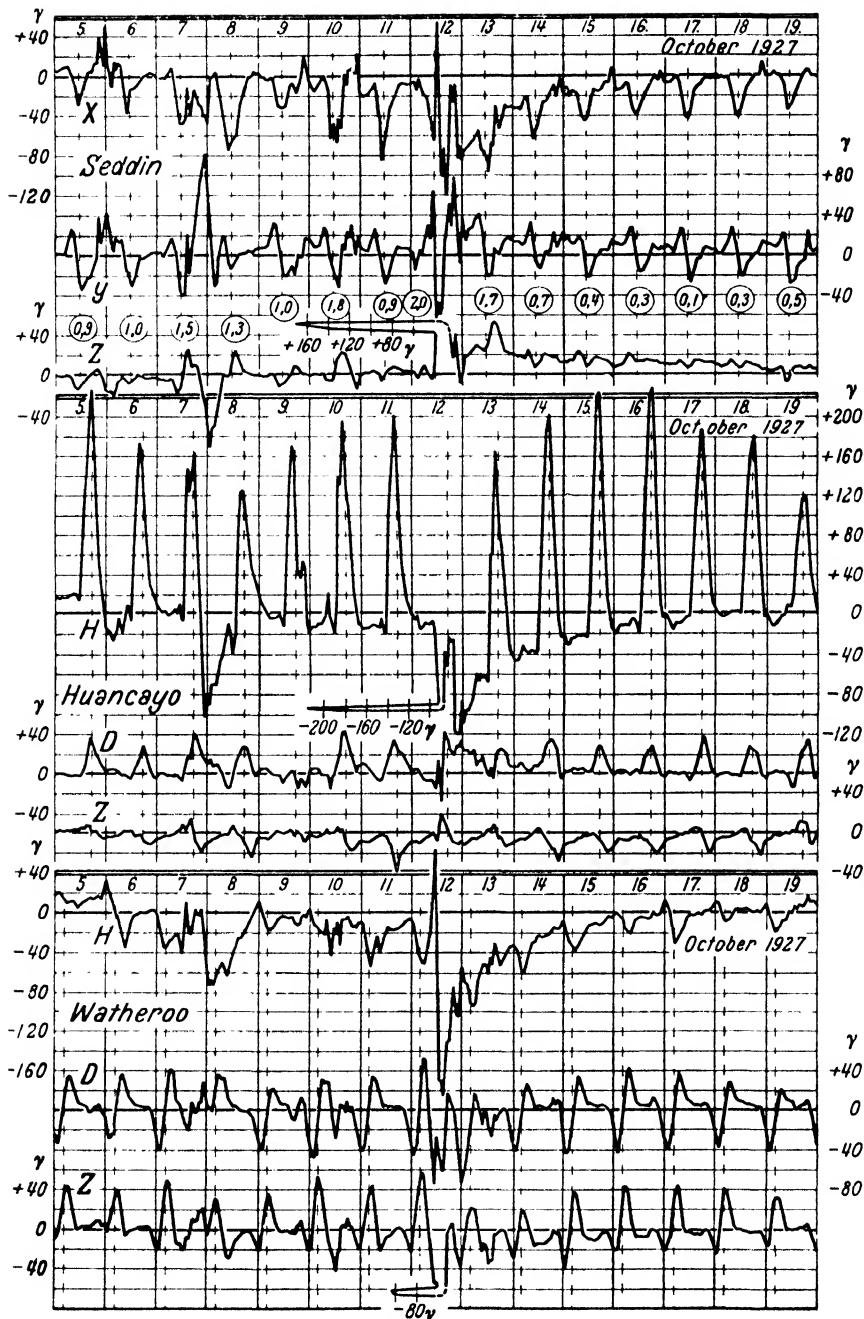
FIG. 3 A. Consecutive simultaneous hourly means of the force components at Seddin (near Potsdam), Huancayo (Peru), and Watheroo (Western Australia) (continued in Fig. 3 B; see description, p. 205)
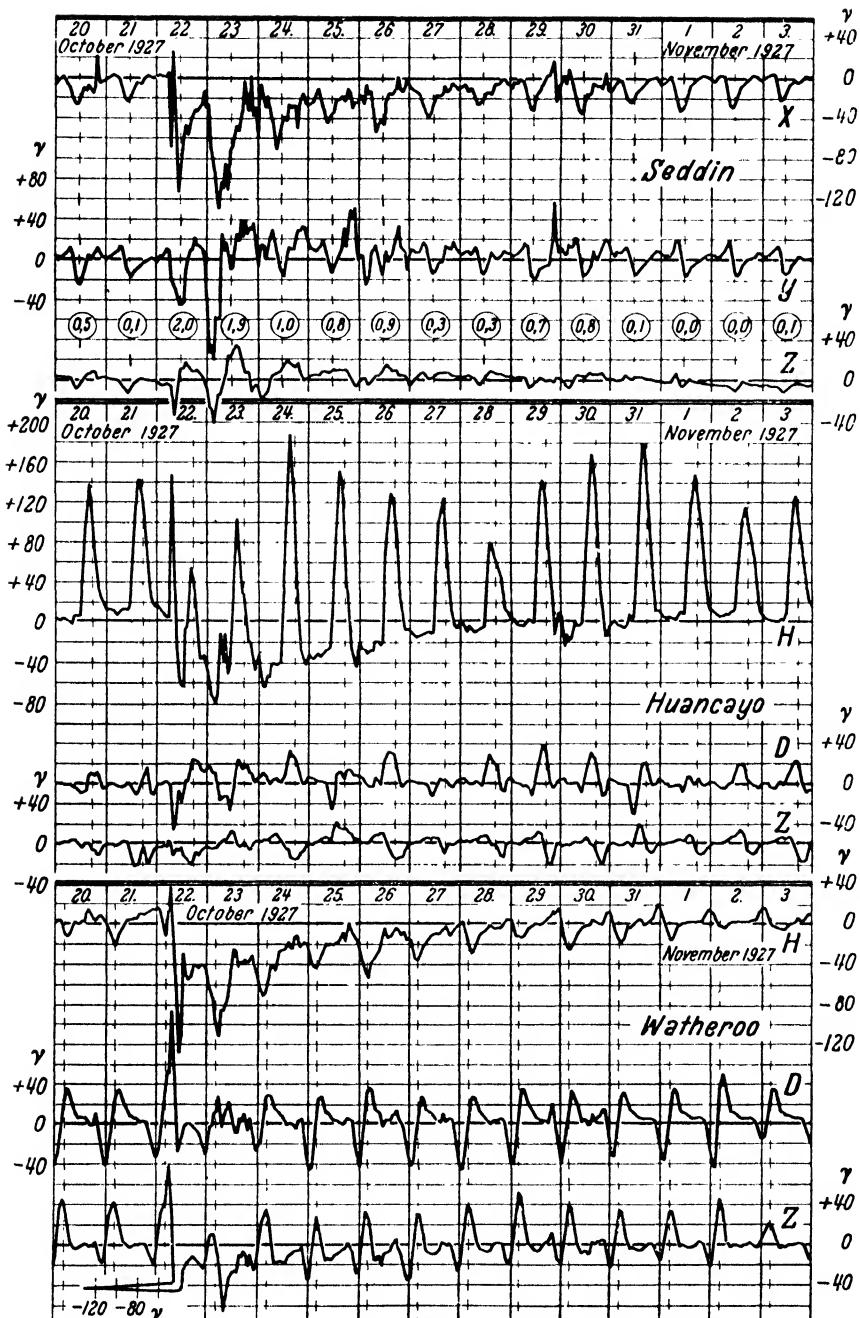
FIG. 3 B. Consecutive simultaneous hourly means of the force components at Seddin (near Potsdam), Huancayo (Peru), and Watheroo (Western Australia) (continued from Fig. 3 A; see description, p. 205)

Peru), and one in the southern hemisphere (Watheroo, Australia). For the two latter observatories, H and D (positive towards the east) had to be taken instead of X and Y, but this does not matter for this pur-
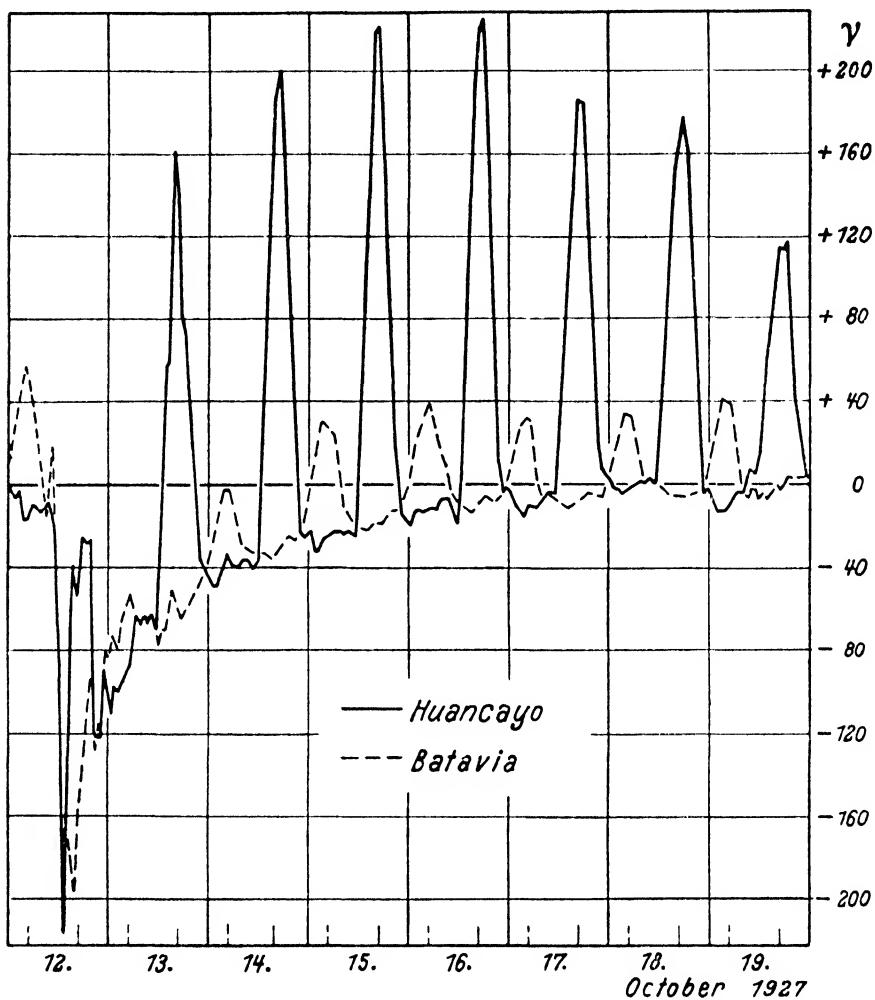
Fig. 4. Consecutive simultaneous hourly means of horizontal intensity at Huancayo (Peru) and Batavia (Java)

pose. The epochs of local noon are again indicated by broken vertical lines. It is striking how similarly at different stations the element X, or H, is affected by storms, while the effect on Z is negligible at the equator, and S_D in Z is of opposite sign in the northern and southern hemispheres.

The daily international magnetic character-figures (§ 2) have been indicated in the upper part of Fig. 3, enclosed in circles between the graphs showing Y and Z for Seddin.

The approximate zero-levels are:

	X or H	Y or D	Z
Seddin	$+18,400\,\gamma$	$-2,000\,\gamma$	$+43,000\,\gamma$
Huancayo	$+29,600\,\gamma$	$+5,500\,\gamma$	$+700\,\gamma$
Watheroo	$+24,700\,\gamma$	$-1,800\,\gamma$	$-51,000\,\gamma$

Finally, Fig. 4 shows, for 1927 October 12–19, the simultaneous hourly means of the horizontal intensity at two equatorial observatories differing in longitude by approximately 180°, namely, Huancayo and Batavia. The zero-level corresponds to $29,622\,\gamma$ for Huancayo and $36,836\,\gamma$ for Batavia. The night values for both observatories follow a strikingly continuous curve; no less striking is the enormous difference between the amplitudes of S_q at the two stations (7.10). Fig. 3 shows the Seddin and Watheroo curves also, for these days.

6.2. Additional notes. The International Association of Terrestrial Magnetism and Electricity, at its Washington Assembly, 1939, decided that its scheme for the provision of numerical character figures (11.4, p. 361) should, for a trial period 1940–2, be replaced by a scheme for the provision of three-hourly range indices (K), to characterize the variation in the degree of irregular magnetic activity throughout each Greenwich day, for the intervals 0–3, 3–6,..., 21–24 hours G.M.T. A committee was appointed to organize the provision of these indices, with special regard to speedy publication. The scheme is a development of the plan adopted by the Potsdam Geophysical Institute for the provision of 'Kennziffern' [11.44]. It will give an international characterization of the variation of irregular magnetic activity, more detailed than that afforded by the daily character figures, but less detailed and voluminous than by hourly measures (cf. 11.5, p. 362). Proposals and provisional values of K (integers ranging from 0 to 9) for the first half of 1938, based on eight well-distributed observatories, have since been published [11.44 a].

VII

THE SOLAR DAILY VARIATION ON QUIET DAYS, S_q

7.1. Geographical distribution and seasonal changes. The quiet-day solar daily variation (S_q) will in general be supposed derived from five quiet days per month (6.2). Even these, however, are often appreciably disturbed in high latitudes, where the intensity of disturbance relative to S_q is much greater than elsewhere. Hence for high-latitude stations S_q is best derived from exceptionally quiet days, chosen as having unusually low international character-figures, or very small daily ranges at the stations themselves.

The distribution of the S_q field depends on two principal coordinates, the latitude l or north polar distance θ, and the local-time longitude λ' relative to the meridian through the sun (that is, the local time t expressed in angular measure). The similarity of the quiet-day daily variations at different observatories in the same latitude is illustrated by Fig. 1. There are, however, cases in which the magnitude of S_q differs appreciably (§ 10) for stations in different longitudes, even near the equator.

Fig. 2 illustrates the course of the variations in the elements X, Y, Z, and I at the equinoxes (when the northern and southern hemispheres are similarly related to the sun). This is a somewhat idealized and smoothed summary of the observations, based on the analysis by Chapman [1].

It appears from the figure that the variations in Y and Z are of opposite sign north and south of the equator, while those in X are of the same sign. Over the range of latitude to which Fig. 2 refers, on either side of the equator, the variation in Y is fairly constant in type and range, except near the equator; in Z the type is nearly constant, but the range decreases as the latitude increases beyond about 50°. In X there is a reversal of type, approximately at the latitudes $\pm 30°$.

S_q undergoes a considerable annual variation, affecting especially the amplitude; but the type is also affected, though to a less extent, except near the latitudes of reversal of type. This is illustrated in Fig. 3 for Bombay.

7.2. Vector diagrams. The diagrams representing S_q by daily inequalities indicate that the daily *changes* are much greater during the hours of sunlight than during those of darkness. This can be illustrated still more clearly by means of vector diagrams. The most complete

LOCAL TIME   LOCAL TIME

EAST DECLINATION·   HORIZONTAL INTENSITY
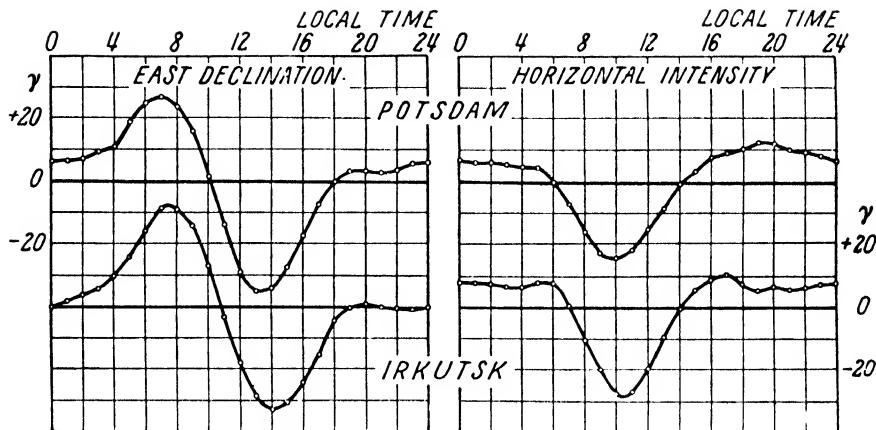
POTSDAM

IRKUTSK

FIG. 1. The solar daily variation in declination and horizontal intensity, in the year 1891, at two stations in equal geographical latitudes: Potsdam (52° 23′ N.) and Irkutsk (52° 16′ N.). (After Schmidt [G 52])

Northcomponent X   Eastcomponent Y   Verticalcomponent Z   Inclination I
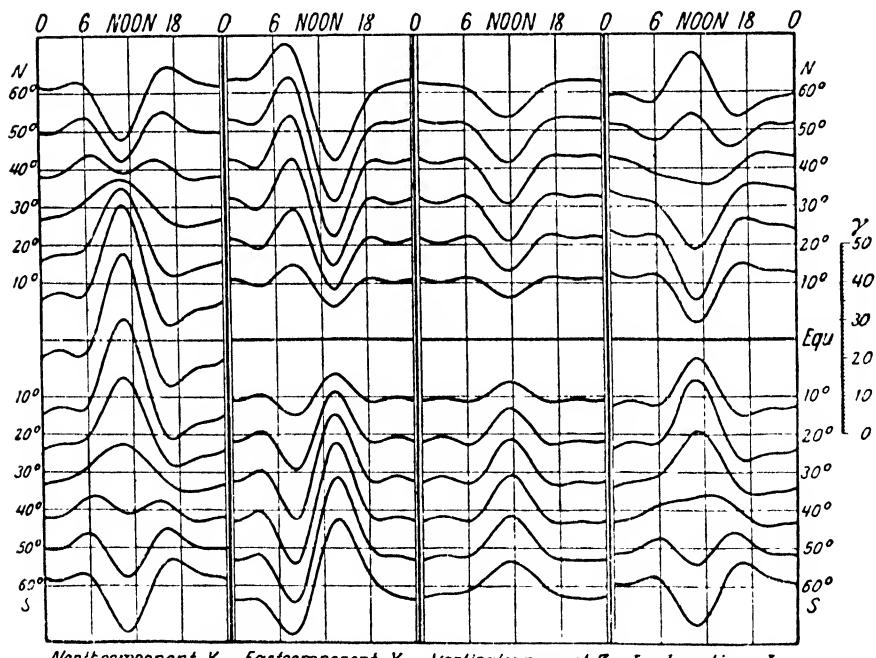
FIG. 2. The solar daily variation of X, Y, Z, and I, at latitudes 10° apart, at the equinoxes, in the sunspot-minimum year 1902
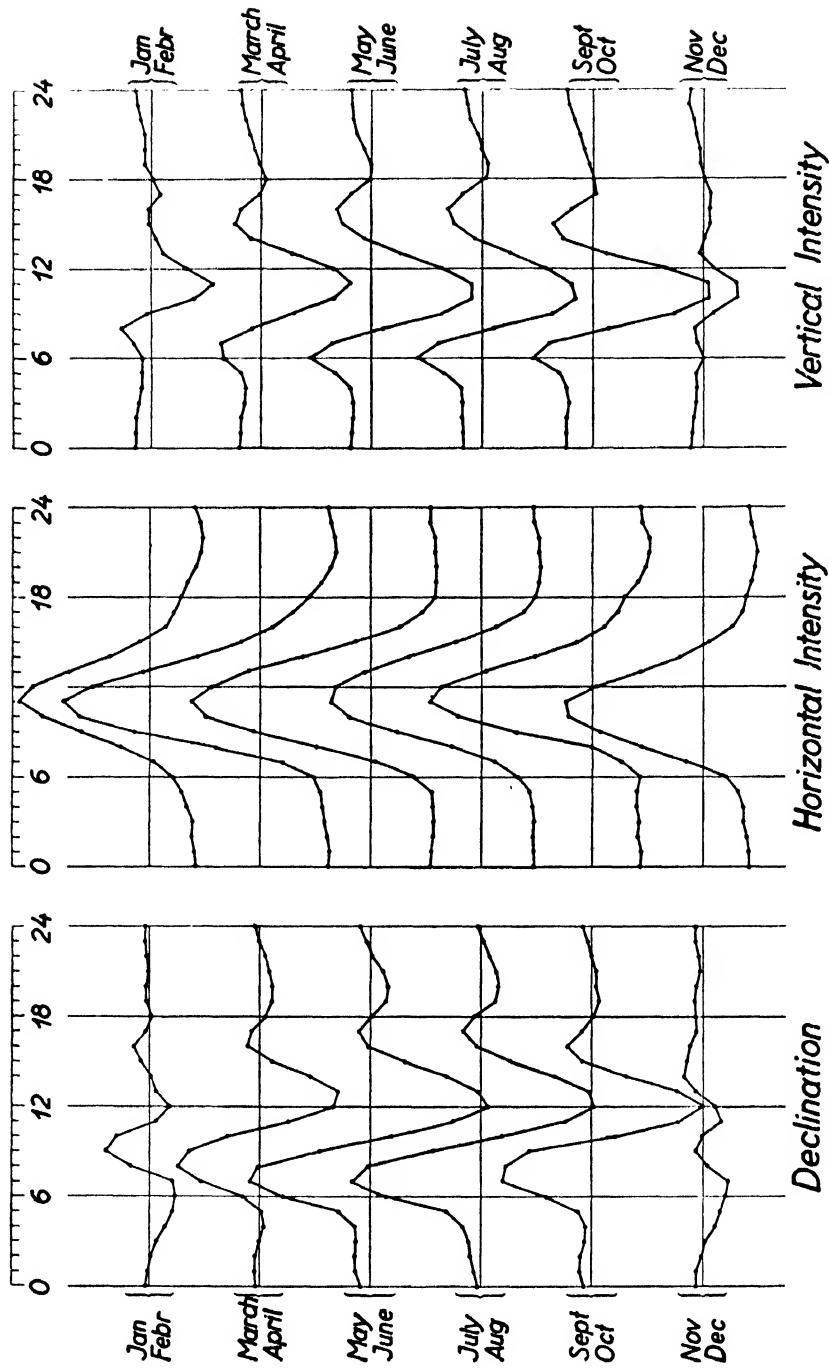
FIG. 3. The solar daily variation on quiet days at Bombay. Bi-monthly means, 1894–1904. Local mean time. Scale: The vertical distance between the zero-lines corresponds to 20γ for H and Z, and to 2 minutes of arc (about 21γ) for D

representation of the S_q change in magnetic force at a single station would be afforded by a model representing the curve traced out (in three dimensions) by a vector (with a fixed origin), equal to the difference (corrected for the non-cyclic variation) between the force at each
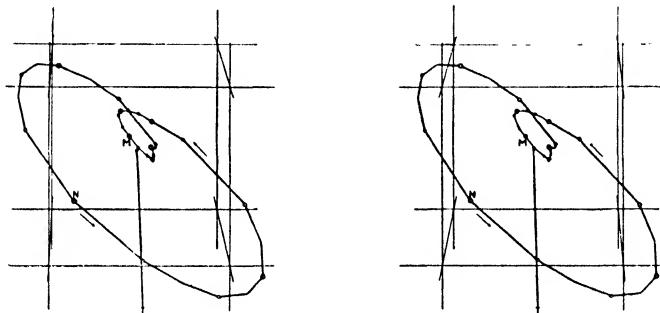
FIG. 4

(See below.) May to August: edge of cube 20 γ
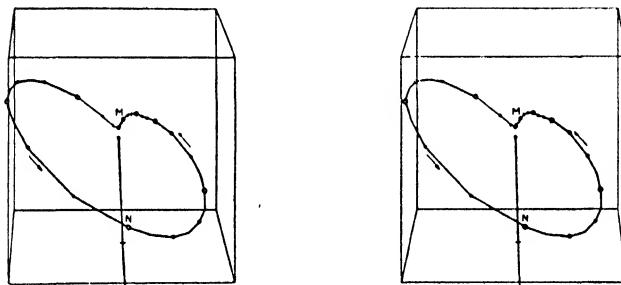
FIG. 5

(See below.) November to February: edge of cube 60 γ

FIGS. 4 and 5. Two stereograms showing the daily magnetic variations on quiet days at Watheroo, Australia, in southern winter (May to August) and southern summer (November to February). Apparent local time, N = noon, M = midnight. The mean value for the whole day is shown in the centre of the cube, near M. The direction of the average field vector is indicated by the line which comes from the south and from below, ending in the centre of the cube. The length of this line between the end and the horizontal mark is 1/4000 (in Fig. 4) or 1/2000 (in Fig. 5) of the total field-intensity [2]. Looking north

instant and its mean for the day. On paper it is not possible properly to represent such a model, except by means of stereograms (Figs. 4, 5); the projection on the horizontal plane, however, as in Fig. 7, sufficiently indicates the feature referred to above. It corresponds to the curve traced out by a vector with a fixed origin, representing in magnitude and direction the horizontal component of the force; it combines the information afforded by the daily inequalities in H and D. Similar

diagrams can be drawn to illustrate, by projections on vertical planes, the joint variations in Z and any one of D, H, X, and Y (Fig. 6).

The vector diagrams also indicate in a striking way the annual variation of S_q in any two elements at a single station. Fig. 7 illustrates this for Greenwich [3], during the months of June (curves I, II) and
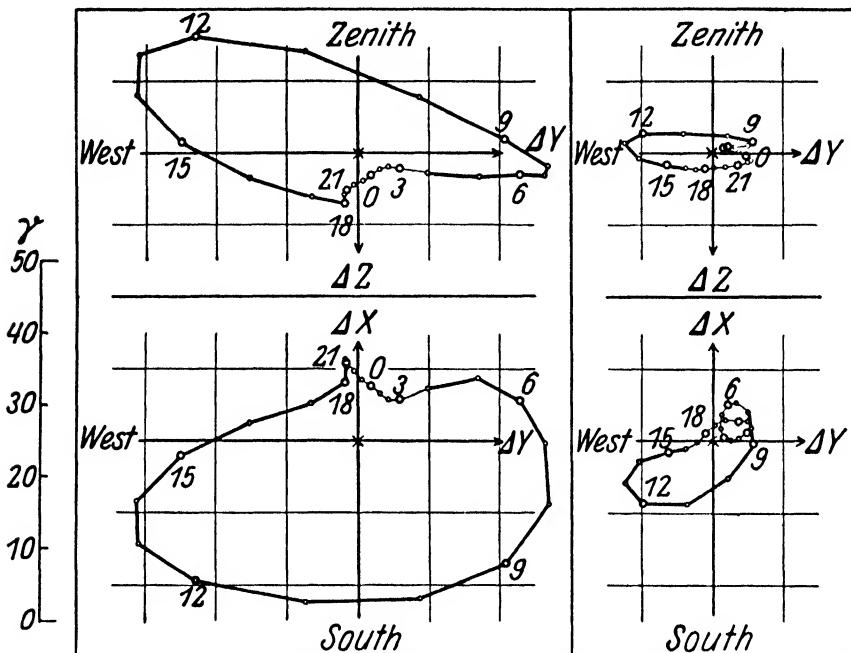
FIG. 6. The solar daily variation at Potsdam, quiet days, 1911–20, June solstice (left) and December solstice (right). Projections of the daily movement of the end-point of the magnetic field-vector on an east–west vertical plane (above) and on the horizontal plane (below)

December (III, IV). The December curves are much smaller than those for June. The portions of the curves traversed during the day are indicated by thicker lines than those traversed during the night; the days are, of course, of very different length at Greenwich at the two seasons. The variation of the rate of change in the horizontal component of force throughout the day and night is well indicated.

Fig. 8 gives horizontal vector diagrams for several stations for the months June and December. It indicates that in extra-tropical latitudes the range of the daily variation in each element is greatest in the summer and least in winter, so that at the June and December solstices the distribution of intensity of the S_q field shows opposed contrasts between the north and south hemispheres.
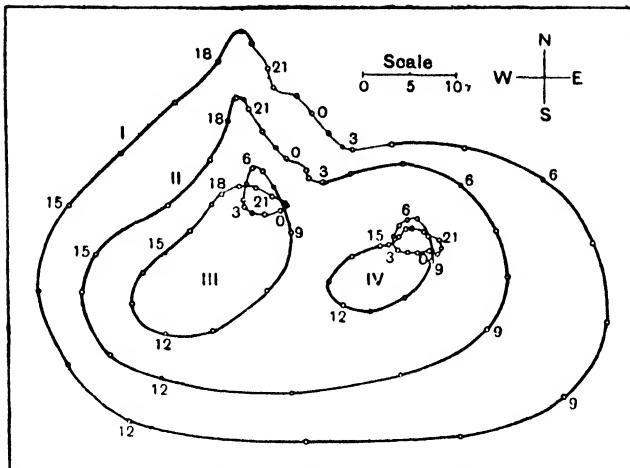
FIG. 7. Quiet-day vector diagrams of the daily variation of
magnetic force in the horizontal plane at Greenwich, 1889–1914,
for June and December, sunspot-maximum years (I and III) and
minimum years (II and IV)
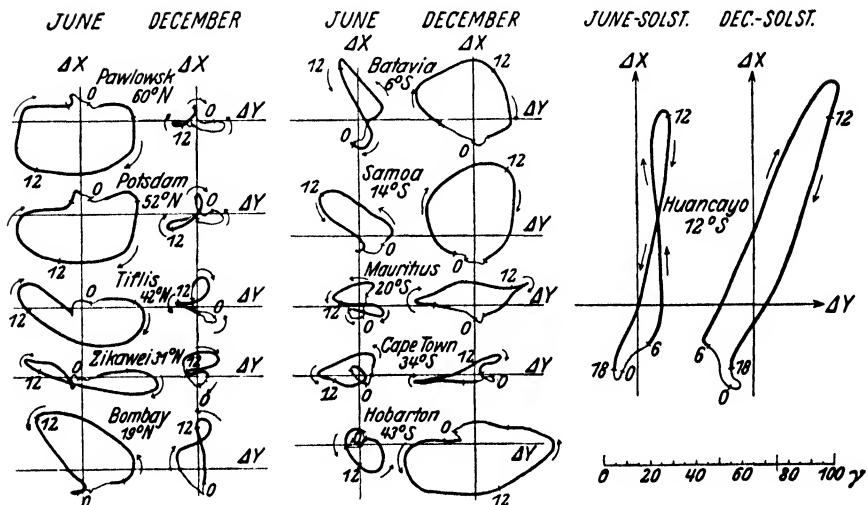
FIG. 8. Horizontal vector diagrams for the solar daily magnetic variations in June and
December at eleven stations, sunspot-minimum years. 0 = midnight, 12 = noon, local
time. The movement between sunrise and sunset is shown by a heavy line

7.3. The increase of amplitude of S_q with the sunspot number.

The two curves for either month in Fig. 7 illustrate the variation of S_q with the condition of the sun's surface; the curves I and III refer to sunspot-maximum years and the others to sunspot-minimum years. The principal difference between the two epochs is a magnification of the S_q curves at sunspot-maximum, though there are also some minor changes of shape indicating a slight variation in the type or distribution of the S_q field. The results from other observatories in the middle belt of the earth, and for Z as well as for X and Y, show a similar dependence on the solar cycle.

The variation of intensity of the S_q field from year to year, in general correspondence with the annual mean sunspot numbers, can be well represented by considering the ranges R'' of the annual or monthly mean daily inequalities of one or more magnetic elements, or alternatively the summed ranges r of the annual or monthly mean daily inequalities; r denotes the sum of the 24-hourly values of the inequality (that is, the deviations from the daily mean), taken without regard to sign. When the inequalities are derived from all days in a month or year they do not strictly represent S_q alone, since some of the disturbance changes are also included; but at stations such as Greenwich, or others in lower latitudes, the disturbance effects largely average out and only slightly affect the range R'' or r of the mean inequalities. The changes of range in the course of the solar cycle are far larger than these small differences, so that it matters little for this purpose whether the inequality ranges are derived from all or only from quiet days. The effect of disturbance would, however, appear to a serious extent in the average of the daily ranges R' (p. 62) of the hourly inequalities for each single day (since the maximum or minimum does not always occur at the same hour), and even more in the average of the absolute daily range R (the difference between the absolute maximum and minimum for each single day), whether for quiet or all days.

Fig. 9, due to Moos [4], shows the variations, over a long period of years, of the monthly mean sunspot numbers s and of the monthly mean summed ranges r in H, D, and Z for Bombay, the values in each case having been smoothed by Wolf's process of taking the mean of two overlapping series of twelve months (for instance, twelve-monthly means from January to December 1890, and from February 1890 to January 1891, are averaged to give the smoothed value ascribed to the month of July 1890). Other stations show generally similar results. It appears from the figure that at Bombay the correspondence with the

sunspot number is closest in the case of H, and least close in the case of Z. The sunspot period is shown very clearly in the H curve, in which, however, there are smaller fluctuations without a counterpart in the solar curves. The curves for r_H show the same inequality as that in s, between the intervals of rise and decline, and the rise to maximum is specially rapid in both the curves for the cycle 1867–78.
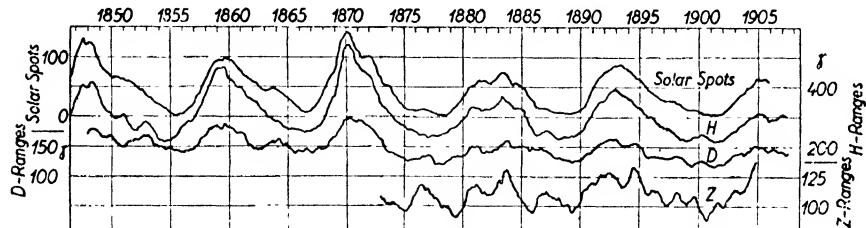
FIG. 9. Relative sunspot numbers and hourly departure sums for the magnetic elements at Bombay. The hourly departure sum is the sum, regardless of sign, of the 24 ordinates of the daily variation curve showing the hourly departures from the daily mean. (After Moos)

Fig. 10, due to Ellis [5], gives similar smoothed curves for the D and H ranges R'' for Greenwich, 1841–77; Fig. 12, due to Maurain [11.38], shows corresponding curves for D, H, and Z for Paris, 1883–1923. At these stations perhaps the D curve is the one most closely parallel to the s curve. Fig. 11 gives unsmoothed curves of s and R'' (in H and D) for Greenwich, for the separate months; the annual variation of R'' in D and H has been eliminated, however, by subtracting from each monthly value the mean value of R'' for that month over the whole period 1841–77. The curve shows that both s and the intensity of S$_q$, as indicated by the measure R'' for Greenwich D and H, are subject to large variations superposed on the more gradual ones which follow the 11-year period. Some of these variations, taking place at short intervals, are comparable in amplitude with the 11-year variation, but though there are some instances of correspondence between those in s and those in S$_q$, there are also numerous cases of disagreement.

Fig. 13 indicates that the intensity of S$_q$ may change considerably even on adjacent days of a single month. It shows the course of H at Bombay during two months, one of which, June 1900, was particularly quiet even for the epoch of sunspot-minimum (the mean s for 1900 was 9·5, and for 1901, the actual minimum year, 2·7): the other month, July 1894, was a typically disturbed month near sunspot-maximum. In June 1900 every day was classed as quiet except June 27, 28, and 29, which were classed as slightly disturbed; but not only did the daily
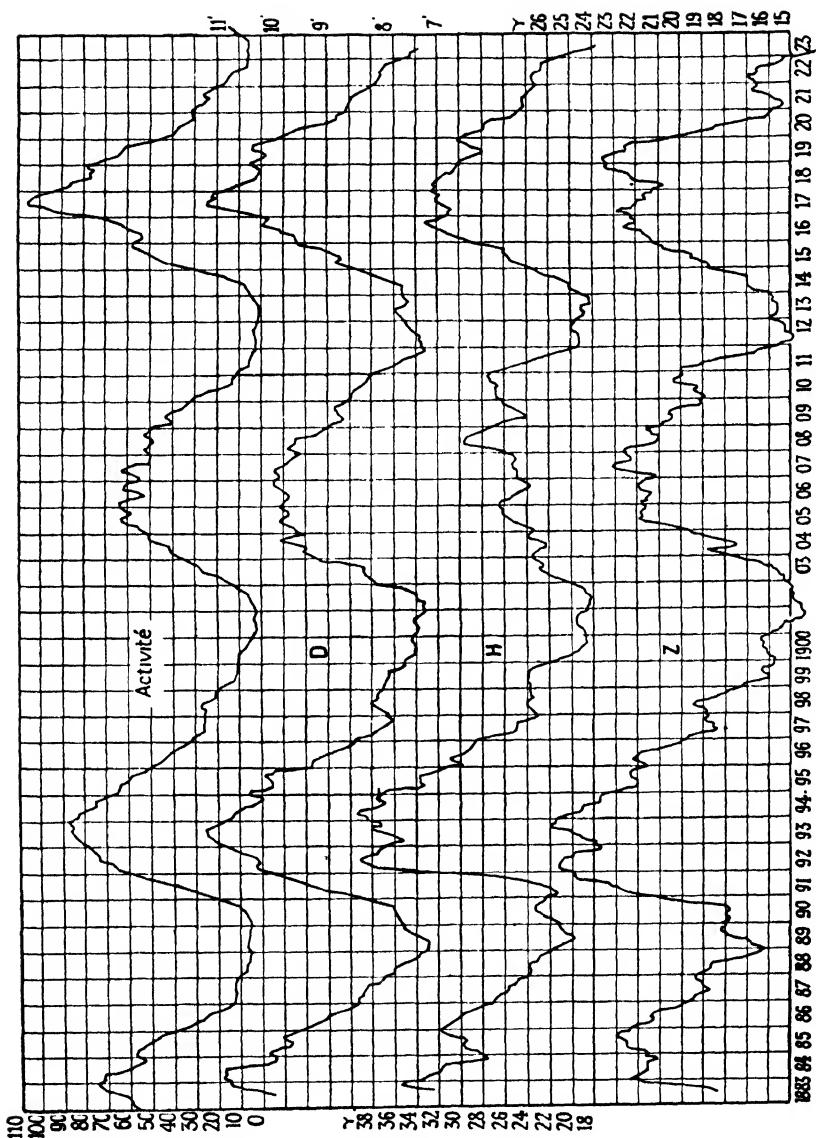
Fig. 12. Smoothed monthly means of relative sunspot numbers (top), and of the daily ranges in declination, and in horizontal and vertical intensity, at Paris, 1883–1923 (after Maurain)

mean fluctuate, as shown by the curve of hourly values smoothed by taking means of successive 24-hourly means, but also the variation about the mean altered greatly in range from day to day—for example, compare the days June 3 and 12 with the days June 8, 14, 15, and 23.
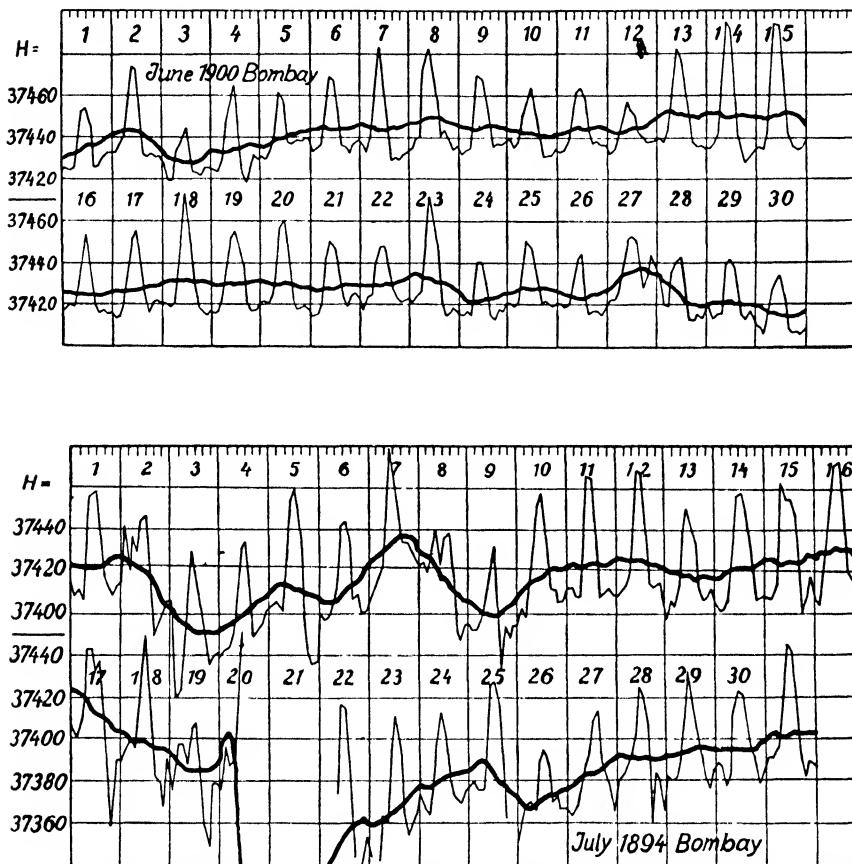
FIG. 13. Hourly values of the horizontal force, and their running 24-hourly means, at Bombay, June 1900 and July 1894 (after Moos)

The type of the variation, however, was substantially the same throughout, except for the great enlargement of the night maximum between the days June 27 and 28. On the other hand, even in a disturbed month such as July 1894, several quiet days, such as July 10–16, may show closely similar daily variations.

Similar variations of S$_q$ are found at other observatories. For example, the two consecutive days 1923 September 21, 22 both received character-figures 0·0, but the individual daily ranges R' in the inequalities (from

hourly means) of D, H, and Z at Greenwich on these days were very different in D and H, though not in Z; they were as follows:

	D	H	Z
Sept. 21	5·4′	24γ	15γ
Sept. 22	9·5′	43γ	15γ

Many similar instances might be cited. The day-to-day variability of S_q is further considered in § 6.

7.4. Wolf's suggested linear relationship. If the curves of R'' (or r) for S_q were exactly similar to those for the sunspot numbers, a linear relationship would hold between the two quantities.

$$R'' = a + bs = a(1 + ms), \tag{1}$$

where $m = b/a$. Wolf [7, 8] was the first to examine whether such a formula represented the observations; he found that for D at Munich (1835–50) and Prague (1840–75) the agreement was fairly good. Evidently a represents the ideal range R'' in a year of no sunspots, while $10^4 m$ represents the percentage increase of range corresponding to an increase of 100 in the sunspot number: the constant m, or $10^4 m$, is of interest as indicating to what extent the different elements at different observatories are affected by sunspot changes.

The following are some typical values of $10^4 m$. The first two are due to Wolf [7, 8], the next to Rajna [10] (based on data from Oslo, Prague, Greenwich, Vienna, and Milan), and the remainder to Chree. Some are from all-day inequalities (a) and some from quiet days (q) only, so that the values given are not strictly comparable.

TABLE 1

Observatory	Lat.	Period	D	H	Z
(a) Munich	48	1835–50	81	..	..
(a) Prague	50	1840–75	78	..	..
(a) Milan	45	1836–94	47	..	..
(q) Pavlovsk	60	1890–1900	68	94	63
(a) Ekaterinburg	57	..	65	109	137
(a) Irkutsk	52	1890–1900	74	104	109
(q) Kew	51	1889–99	66	104	54
(a) Zikawei	31	..	69	..	..
(q) Bombay	19	1894–1901	28	89	37
(a) Batavia	−6	1887–98	72	71	51
(a) Mauritius	−20	1886–90	40	64	58

These suggest that on the whole the Z-component of the S_q field is increased in a smaller ratio than D and H at sunspot-maximum. It should be noted, however, that Wolf's formula does not fit the observa-

tions with great accuracy, and that different values of m are obtained in different sunspot cycles, even for the same element and station.

Chree examined how far m changes with the season. He found that at Kew and some other northern stations (though not in Greenwich D) it is larger in winter than in summer; at tropical stations its seasonal changes are rather uncertain. He also examined how far a linear expression such as $a+bs$ applies to the year-to-year variations of the Fourier coefficients of the S_q inequalities for different elements and stations [15–17, also G 15].

Schmidt [12, 13] applied Wolf's formula to the individual hourly departures in the (all-day) inequalities for X, Y, Z, H, D, I, and F at twelve northern and southern stations, determining, for each element and hour at each station, a separate a and b. For the same element and station the coefficients b are not constant, even in sign, throughout the day, showing that not only the magnitude but also the *type* of S_q (or rather of S_a, for all days) changes somewhat in the course of the sunspot cycle (Fig. 14). By means of his values of a and b he calculated the hourly inequalities for the twelve stations corresponding to a mean sunspot epoch ($s = 50$). In this investigation he considered the three seasons separately, and also the individual months June and December.

7.5. The harmonic analysis of S_q. The daily inequalities of the magnetic elements can be specified also by their Fourier constants (2.15 and Ch. XVI), usually, with sufficient approximation, for the first four harmonics. The table on p. 227, taken from Schmidt's *Archiv* [13], gives values of c_n and ϵ_n for $n = 1$, 2, and 3 for a small number of northern and southern stations at the three seasons: the inequalities analysed refer to *all* days and to mean sunspot years ($s = 50$). The table indicates in its own way some of the features already noted in Fig. 8; for example, at Bombay and Singapore, in magnetic latitudes $9°$, $-10°$, the amplitudes and phases of the components of X are similar, but the phases are opposite (that is, about $180°$ different) from those at Pavlovsk. Tiflis and Capetown are not far from the latitudes of reversal of S_q in X, and the amplitudes are small and the phase-angles variable. The reversal of the phase of S_q in Z between Bombay and Singapore is shown in all three phase-angles; in Y the reversal is much less definite, and as regards this element the two stations behave somewhat as if they were both northern in the northern summer (June solstice) and both southern in the southern summer (December solstice).

The fourth component also shows a considerable degree of uniformity in its distribution; at many stations its amplitude (in X, Y, and Z) is

greatest near the equinoxes, with intervening minima near the solstices. The changes throughout the year in the amplitudes and phase-angles of this and the other components can best be followed by considering the values derived from inequalities for the twelve separate calendar months.
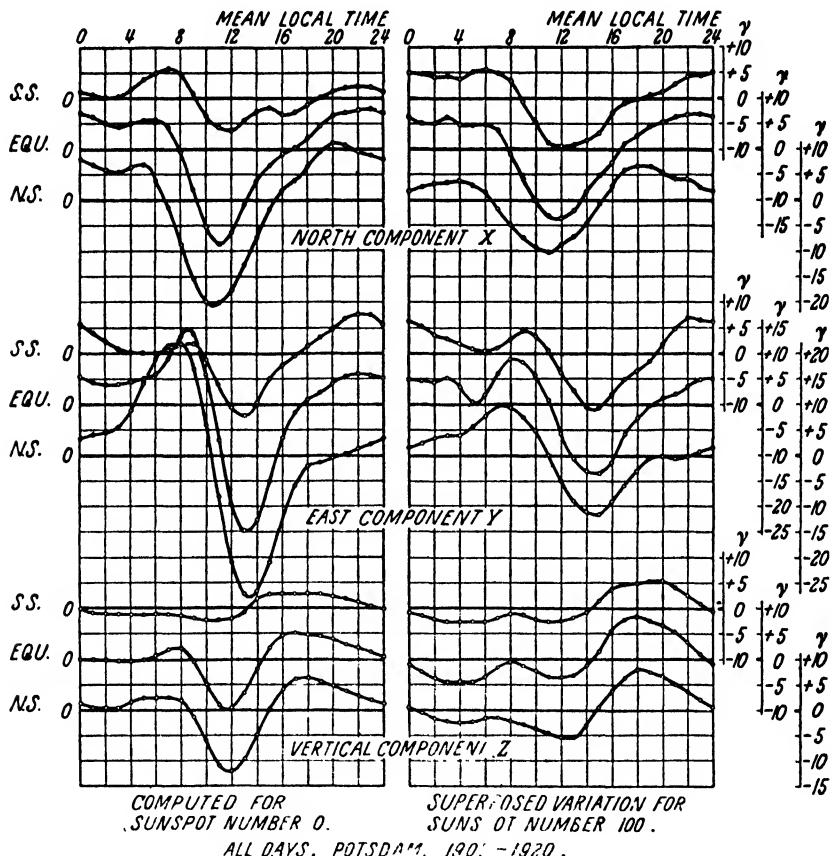
FIG. 14. The solar daily variations of the magnetic force components at Potsdam, 1900–20, all days, showing the computed variation for zero sunspot number (left), and the superposed variation for an increase of the relative sunspot number to 100 (right). Four-monthly groups, S S. = southern (December) solstice; N.S. = northern (June) solstice; Equ. = equinoxes (after Schmidt)

The Fourier constants in Table 2 are derived from all-day inequalities, but except in high latitudes the values derived from quiet days are not very different. The differences are, however, systematic, and particularly affect the first component.

The Fourier coefficients of S_q for the elements X, Y, and Z at a number of stations afford the most convenient material for the spherical

TABLE 2

Harmonic analysis of the solar daily variation, all days, mean relative sunspot number 50. The three lines for each observatory are averages for the four-monthly seasons, June solstice (j), equinoxes (e), December solstice (d). (After Schmidt)

Observatory		North component X						East component Y						Vertical component Z					
		c_1	c_2	c_3	ε_1	ε_2	ε_3	c_1	c_2	c_3	ε_1	ε_2	ε_3	c_1	c_2	c_3	ε_1	ε_2	ε_3
		γ	γ	γ				γ	γ	γ				γ	γ	γ			
Pavlovsk	j	17·3	8·3	3·0	113°	290°	148°	19·6	12·9	4·5	22°	225°	40°	3·9	3·5	0·9	124°	264°	84°
	e	14·0	7·0	4·0	99°	273°	130°	11·9	9·6	5·5	24°	206°	33°	1·9	2·7	1·3	133°	257°	81°
	d	3·9	3·6	2·1	98°	251°	127°	5·0	3·4	2·2	46°	185°	34°	2·0	0·7	0·3	191°	286°	162°
Tiflis	j	5·3	7·1	6·0	129°	8°	195°	26·0	19·9	9·2	31°	222°	62°	6·5	5·4	2·6	89°	264°	94°
	e	3·5	4·6	4·5	77°	16°	194°	17·6	15·8	10·7	38°	208°	50°	4·8	4·8	3·5	98°	263°	96°
	d	5·1	1·0	10·0	4°	204°	143°	7·7	6·4	4·8	72°	176°	43°	3·2	1·8	1·3	136°	264°	99°
Bombay	j	20·0	10·7	2·7	280°	112°	304°	10·4	10·9	8·1	31°	255°	94°	6·3	6·7	5·3	90°	313°	152°
	e	19·6	10·7	4·8	277°	111°	317°	6·2	6·8	7·2	32°	246°	86°	6·6	5·8	5·5	90°	316°	154°
	d	16·2	7·9	2·9	284°	120°	326°	1·3	1·3	1·9	184°	113°	51°	3·1	1·1	1·8	72°	297°	139°
Singapore	j	20·8	9·7	3·0	285°	118°	317°	4·4	5·7	3·0	359°	280°	107°	8·5	3·5	0·6	273°	84°	348°
	e	23·3	11·0	5·7	284°	110°	327°	2·8	4·4	1·8	169°	337°	189°	6·4	4·2	2·3	281°	100°	308°
	d	20·5	8·2	4·1	285°	109°	322°	10·5	8·0	3·3	189°	15°	233°	5·9	4·0	2·2	279°	95°	288°
Capetown	j	5·4	1·9	1·8	305°	222°	45°	4·4	4·6	4·2	77°	304°	139°	2·8	3·1	4·9	314°	355°	175°
	e	2·6	3·4	1·7	337°	356°	139°	8·8	7·9	6·2	143°	342°	185°	6·7	9·9	7·5	263°	52°	227°
	d	2·7	3·8	2·8	168°	338°	150°	13·4	9·0	5·5	153°	15°	213°	8·7	9·0	5·5	238°	59°	241°

harmonic analysis of the S_q field, after the method applied by Gauss to the main field of the earth. The method was first applied to the field of the solar daily variation by Schuster. The subject is discussed in detail, along with the similar analysis of the field of the lunar daily magnetic variation L, in Chapters XIX and XX. The solar daily variation in high latitudes is discussed in 9.7 and 9.8.

7.6. Representation of S_q by overhead current-systems. The spherical harmonic analysis, as described later in Chapters XVII and XX, leads to a determination of the intensity and distribution of the system of horizontal electric currents in the upper atmosphere (the ionosphere) that could produce the S_q field. This current-system is indicated by world maps in Figs. 15 and 16 (cf. p. 696), which relate respectively to the equinox and the June solstice (northern summer). The curved lines are those along and between which the currents flow in the direction indicated; a current of 10,000 amperes flows between each pair of adjacent lines. The current-system consists of four sets of ovals, two in each northern or southern hemisphere; in either such hemisphere, one set is situated in the day-lit portion, while the other extends over the night area. The current-systems occupy a constant position in relation to the sun, or to local time; hence, as the earth revolves, their geographical position continually changes, moving westwards with the sun.

These diagrams may be regarded merely as convenient synthetic representations, in a particular way, of the distribution of the S_q field, without implying any special theory of the origin of the field—without even implying that it is due to overhead *currents*. They are particularly valuable in giving a simple world-view of the order of magnitude of the seasonal variation of the field, by comparison of the current-systems in the northern and southern hemispheres of Fig. 16 with each other or with those of the equinoctial Fig. 15. An approximate method [9.19] of deriving the quantitative features of such current-systems from the magnetic observations of S_q is described in § 7.

It may first be explained, however, how the current-systems account qualitatively for the general nature of the S_q daily variation curves in different elements and for stations in different latitudes. This will be done by considering Figs. 2 and 15 for S_q at the equinoxes; the other seasons can be dealt with in like manner.

At a point below an infinite plane horizontal sheet carrying a uniform rectilinear distribution of current, there is a uniform horizontal magnetic field whose direction, viewed from above the sheet, is perpendicular to
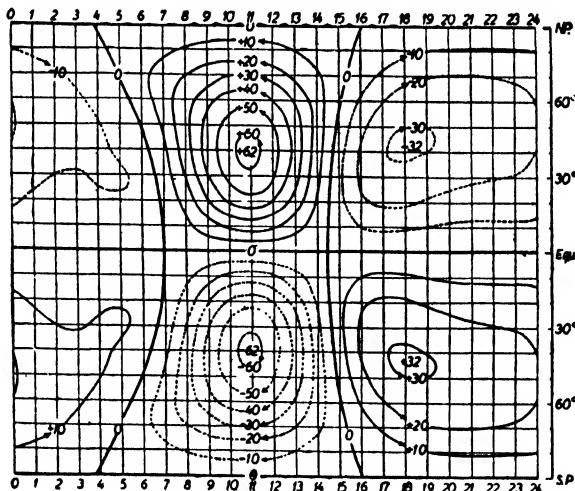
FIG. 15
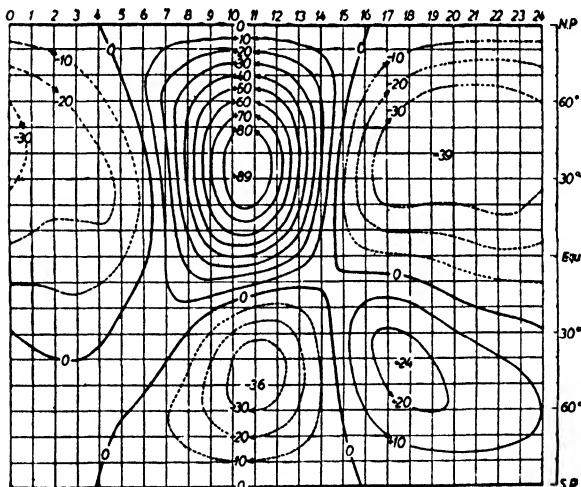
FIG. 16

FIGS. 15 and 16. World maps of the ionospheric current-systems corresponding to the solar daily magnetic variations in the sunspot-minimum year 1902. Top: Equinoxes; below: June solstice. The meridians refer to local time, 12 = noon. Between consecutive stream-lines, 10,000 amperes flow in the direction of the arrows (after Bartels)

that of the uniform current; its sense corresponds to an anti-clockwise rotation of the current-direction, through 90°.

At a point P not far below a non-uniform current-distribution over

a spherical sheet, such as is represented by Figs. 15 and 16, the horizontal component of magnetic force will agree approximately in direction and sense, and even in general in magnitude, with that due to an infinite plane horizontal uniform current-sheet which is tangent to the spherical sheet at the point P' vertically above P, and which has the same intensity and direction as the actual current-flow at P'. For example, in Fig. 15 the horizontal magnetic force due to the sheet will be northwards on the equator between 7^h and $14\frac{1}{2}^h$ local time, and southwards at other local times; it will be greatest, on the equator, at about 11^h, where the current-intensity is relatively large (as indicated by the closeness of the adjacent current-lines); at 7^h and $14\frac{1}{2}^h$, where the current changes direction, its intensity is zero, and there also the horizontal magnetic force will be (approximately) zero. Thus at the equator X will have a maximum at about noon, will pass through zero at about 7^h and $14\frac{1}{2}^h$, and will have a minimum (of less magnitude than the maximum) at about 17^h or 18^h. This is in fair agreement with the daily variation curve for X at the equator in Fig. 2 (complete agreement is, of course, not to be expected, both because of the difference between the spherical current-system and the infinite plane uniform current-sheet, and also because a minor part of the S_q field is produced by a current-system within the earth, as proved in Chapter XX and explained in Chapter XXII).

Consider next the X variation at a station in latitude 60° N. In this latitude the current-direction is not purely eastwards or purely westwards, as at the equator: at most local times it has both an east–west and a north–south component. The direction of the horizontal component of the magnetic force due to the currents, below the current-sheet, being at right angles to the direction of current-flow, will likewise have both an X and a Y component. In considering the sign of the X component, we must have regard to the direction of the east–west component of the current-flow. At 60° latitude this component in Fig. 15 is westwards between $5\frac{1}{2}^h$ and 15^h, so that ΔX is negative (southwards) between these hours, with a minimum near 11^h, where the current-intensity is greatest; the minimum is roughly equal to the maximum at the equator. During the remainder of the day, at 60° N., ΔX is northwards, that is, positive, with a maximum at about 20^h. These indications of the current-system are in fair agreement with the X daily variation for 60° N. shown in Fig. 2.

The X variations at the equator and at latitude 60° N. are thus almost opposite to one another: the transition between the two should,

according to Fig. 15 for the current-system, come at about latitude 40° N., where the current-ovals (on the sunlit hemisphere) shrink to a point C (say). At this point C the magnetic force due to the current-system will be vertical and upwards; hence at 40° N. at 11^h (the local time corresponding to the point C) ΔX and ΔY should both be zero, whereas ΔZ should have a minimum (since Z is positive downwards). All these three inferences from the current-system are in agreement with the X, Y, Z daily variation curves for latitude 40° N. in Fig. 2. The vanishing-point C' of the current-ovals in the night hemisphere is somewhat north of 40°, and is at the local time about 18^h. At 40° latitude the current-flow is westwards, but weak, during the greater part of the night, when ΔX will be negative; it is weak and eastwards throughout most of the period from 3^h to 18^h, during which ΔX should be positive; this agrees fairly well with the X curve for 40° N. in Fig. 2.

As regards the daily variation of Y corresponding to Fig. 15, this should be greatest in the latitudes where the current-flow has the largest north–south component; it should vanish (at the equinoxes) on the equator, and be reversed on crossing the equator, because at corresponding points north and south of the equator the component of current-flow along the meridians is in opposite directions. Since along the meridian 11^h the current-flow is to east or west, ΔY should be zero there, as Fig. 2 shows; it should have a maximum (ΔY eastwards) in the northern hemisphere at about 7^h, when the southward current is most intense, and a minimum (ΔY westwards) at 14^h or 15^h, when the northward current is strongest. During the hours covered by the current-lines bearing negative numbers (that is, in the main, during the night hours) the current-intensity is low, and ΔY should be small; it should be positive except between 11^h and 18^h, when the flow is northwards. These inferences are in fair agreement with Fig. 2.

The value of ΔZ is less readily inferred from the current-diagrams, but it is clear that at the vanishing-points C and C' of the (northern) current-ovals, ΔZ will have its minimum and maximum values: hence the Z daily variation should have its greatest range in latitude about 40°, with a minimum at 11^h and a maximum at about 18^h. It should have the same form at all latitudes north of the equator, but its range should decrease to zero at the pole, and also at the equator, because of the symmetry of the current-system relative to the equator. In the southern hemisphere ΔZ should be reversed, owing to the opposite cyclic direction of the current-flow round the southern vanishing-points

corresponding to C and C'. These inferences agree with the ΔZ curves in Fig. 2.

It should be remembered, however, that both Fig. 2 and Fig. 15 refer to idealized simplified versions of S_q; in particular, S_q does depend somewhat on the longitude as well as the local time, so that the corresponding current-system will not be truly constant as viewed from the sun, nor will it be strictly symmetrical with respect to the equator, even at the equinoxes.

7.7. The approximate determination of the current in overhead circuits. The spherical-sheet current-distributions shown in Figs. 15 and 16, such as could approximately account for S_q, have been derived from the magnetic observations by the rather complicated mathematical process of spherical harmonic analysis, which will be described later (Chs. XVII, XX). Their general nature, however, can be inferred by simple inspection of the daily variation curves for the three magnetic elements at a set of stations suitably distributed in latitude, using considerations of the kind set out in the last section. Fortunately it is possible also to determine approximately the current-intensity and the total current in the various circuits, using a simple method introduced by Chapman [9.19].

Consider the northern current-circuit in Fig. 15 (for the equinoxes), enclosing the point C (in 40° N. at 11^h) already referred to. At its lower border, at the point C_0 on the equator south of C (at 11^h), the current-flow is fairly parallel and uniform, and the magnetic force at a point immediately below will be nearly the same as that due to an infinite horizontal plane uniform current-sheet with the same direction and intensity of current-flow. Let this intensity be denoted by i_0 (eastwards), reckoned in electro-magnetic units of current per cm. breadth in the horizontal direction at right angles to the current-flow. Then the horizontal magnetic force at C_0 will be $2\pi i_0$ (1 (29)), northwards. This, however, is not the observed magnetic intensity ΔX_0 at C_0, because ΔX_0 includes a contribution due to a secondary current-system flowing *within* the earth. The existence of this system is definitely established by spherical harmonic analysis of the S_q field (Ch. XX), and the fraction f of ΔX_0 due to the atmospheric current-system is shown to be approximately 0·6. This reference to the spherical harmonic analysis is the only point at which the present method requires the support of complicated mathematics. Thus, equating the fraction f of the observed X variation at C_0 to the magnetic force $2\pi i_0$, we have

$$f\Delta X_0 = 2\pi i_0, \tag{2}$$

or
$$i_0 = \frac{f\Delta X_0}{2\pi} \text{ e.m.u./cm.}$$

$$= \frac{10f\Delta X_0}{2\pi} \text{ amperes/cm.} \tag{3}$$

Taking ΔX_0 as 20γ (cf. the curve for X at the equator in Fig. 2 for the time 11^h), this gives $i_0 = 1.9 \times 10^{-4}$ amp./cm. or 19 amps./km. or 21,000 amperes per $10°$ range of latitude. This may be compared with the value 23,000 amperes found from spherical harmonic analysis, as indicated in Fig. 15.

In order to get the total current I_{day} flowing round the daytime current-circuit, we need to know the mean current-intensity $\bar{i}$ between the centre of the circuit (at C in $40°$ latitude) and the equator. At C, and at the corresponding point C_1 south of the equator, i must be zero, because there it has no definite direction. If i varies 'parabolically' between C and C_1, with the maximum value i_0 at the equator, then up to the latitude l_0 of C,
$$i = i_0 - al^2,$$
where
$$al_0^2 = i_0.$$
Hence the mean value of i between the equator and C is given by

$$\bar{i} = \frac{1}{l_0}\int_0^{l_0} i\, dl = \frac{1}{l_0}\int_0^{l_0}(i_0 - al^2)\, dl = \frac{1}{l_0}(i_0 l_0 - \tfrac{1}{3}al_0^3)$$

$$= i_0 - \tfrac{1}{3}al_0^2 = \tfrac{2}{3}i_0.$$

Alternatively, if the curve of variation of i between C and C_1 resembles a sine curve, so that $i = i_0\sin\left(\tfrac{1}{2}\pi\dfrac{l_0 - l}{l_0}\right)$, with zero values at $l = l_0$ and $l = -l_0$, and the value i_0 at $l = 0$ (the equator), then it is readily proved that $\bar{i} = 2i_0/\pi$, which does not differ materially from the previous value of $\bar{i} = 2i_0/3$.

Consequently, we take $I_{day} = \bar{i}l_0 = \tfrac{2}{3}i_0 l_0$, if now i_0 is reckoned per degree of latitude; thus reckoned, its value is 2,100 amperes. Hence, taking $l_0 = 40°$,

$$I_{day} = \tfrac{2}{3} \times 2,100 \times 40 = 56,000 \text{ amperes.}$$

This is a reasonably good approximation to the value of 62,000 amperes derived from the spherical harmonic analysis.

The same method may be applied to the night current-circuits, whose centres C' and C_1' lie in approximately the same latitude ($40°$) as C and C_1 and on the 18^h meridians, as is best inferred from the passage of the

east-force variation, ΔY, through its zero value at that hour. Taking the night value of ΔX_0 at the equator as 12γ, we have

$$I_{\text{night}} = (12/20)I_{\text{day}} = 33{,}600 \text{ amperes};$$

the value derived from the spherical harmonic analysis is 32,000.

Across the circle of latitude l between C and C′ (11^h and 18^h) there flows northwards the combined current of the day and night circuits, namely, 89,600 amperes according to our approximate calculations, or 94,000 according to spherical harmonic analysis. This combined value can be inferred independently from the daily variation of the east force Y in latitude l_0. The distance in centimetres between C and C′, which are in latitude $40°$ and 7^h apart in longitude, is $8{\cdot}9 \times 10^8$. The value of the S_q horizontal component in latitude $40°$ at about 14^h, when it is eastwards, can be seen from the daily variation curve for Y (Fig. 2); it is about 16γ, of which we suppose $9{\cdot}6\gamma$ is due to the overhead currents, and equal to $2\pi i_0$. This gives $i_0 = 1{\cdot}5 \times 10^{-4}$ amp./cm., and $\bar{\imath} = 1{\cdot}0 \times 10^{-4}$ amp./cm. Hence $I = 1{\cdot}0 \times 10^{-4} \times 8{\cdot}9 \times 10^8 = 89{,}000$ amperes, in good accord with the value 89,600 based on data for the north component X alone.

The southward current in the northern hemisphere, spread over the twelve night hours 18^h to 6^h, is less intense, and its intensity fades away towards the dawn meridian at 6^h. Its distribution can be approximately inferred from the variation of Y during the night. The departure of Y from the daily mean is almost zero in the $10{\cdot}5$ hours between $17{\cdot}5^h$ and 4^h, and the current-density must be very small during this interval. Over the seven-hour interval 4^h to 11^h the maximum divergence of Y from the mean is 14γ, giving (by comparison with the Y variation of 16γ between 11^h and 18^h) a total southward current of

$$(14/16) \times 89{,}000 = {\cdot}78{,}000 \text{ amperes};$$

the remaining 11,000 amperes must be spread over the long interval $17{\cdot}5^h$ to 4^h.

This approximate method of determining the intensity of a spherical current-system corresponding to a given surface-field, although useful for many purposes, should not supersede the more accurate method of spherical harmonic analysis (Chs. XVII and XX), where this is practicable.

7.8. Day-to-day changes of intensity of S_q. It has already been mentioned in §3 that at some stations (e.g. Bombay and Greenwich) the S_q variations on quiet days vary in intensity from day to day. Chapman and Stagg examined the day-to-day changes of S_q at the two

stations Eskdalemuir and Greenwich [20], and later at other stations. To minimize any influence of magnetic disturbance, they considered only really quiet days, namely, all the days of international character-figure 0·0 and 0·1 during the period 1913–23. The range R in each element was tabulated for all such days; their regular changes in the course of the year and the sunspot cycle were determined and were used to define a 'normal' range R_n for any date in any of the years 1913–23. The actual ranges R usually differed from the 'normal' values R_n for the respective days, and the percentage difference ΔR from the normal, i.e. $100(R-R_n)/R_n$, was tabulated for each day, there being one such ΔR for each element at each of the two observatories. Naturally some of the values of ΔR were positive and others negative; the average numerical value, without regard to sign, was different for different elements and seasons, and ranged from 20 to 30 per cent.; the distribution of the ΔR's was fairly symmetrical about the zero value, and similar in different elements and at the two observatories.

They found that the two values of ΔR for the same *element* at the *two* observatories on the same day were closely correlated, the correlation coefficient (see 11.10) for the north force being 0·77 and for the west force 0·84. For the vertical force the correlation was less, and varied much more with the season than in the case of the north and west components of force; the values of the correlation coefficient for the vertical force, for summer, equinox, winter, and the year, were 0·64, 0·50, 0·25, and 0·47.

The correlation between the values of ΔR on the same day at the *same observatory* but for *different elements* was also examined, and found to be much less close than that between ΔR for the *same element* at the two *different observatories*. The correlation between the values of ΔR for north force and west force was fairly constant throughout the year, the mean yearly value being 0·38 at Eskdalemuir and 0·41 at Greenwich; the corresponding correlations for west force and vertical force were 0·32 at Eskdalemuir and 0·23 at Greenwich. The correlation between north force and vertical force was only very slight, namely, 0·18 at Eskdalemuir and 0·07 at Greenwich. In these inter-correlations at the same observatory the coefficient was found in general to be the same for years of high as for years of low solar activity.

The average *type* of the S$_q$ variation was shown to be the same, at a given season, on (quiet) days of large or small range, so that S$_q$ on such days, on the average, varies only in *intensity*; hence the range R in each element appears to be a true index of the difference between

the S_q variations on different quiet days. This constancy of the average *type* of S_q was shown by determining the mean daily variation for each element, at Eskdalemuir, from three groups of days (each containing approximately the same number of days) classified according to their values of ΔR; one group contained the days of largest positive ΔR, another those of largest negative ΔR, and the third group contained
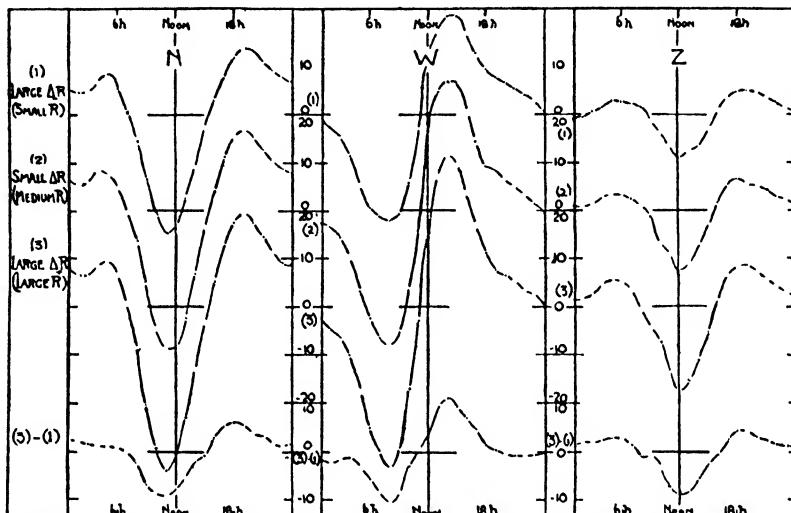
FIG. 17. The daily variation on quiet days, at Eskdalemuir, June–July 1913–23; north, west, and vertical components. Averages for three groups of days having small, medium, and large daily ranges; the difference-curve (third minus first) is also shown

the days on which ΔR had intermediate values; the subdivision was made separately for each element. The 156 days of character 0·0 or 0·1 in the months June and July 1913–23.were thus grouped, and the mean daily hourly inequality for each group was formed. The inequalities are plotted in Figs. 17 and 18; the curves show that the considerable differences of amplitude in the daily inequalities for the three groups are unaccompanied by any appreciable change of type; this is more clearly brought out by the lowest curves, which represent the difference between the curves for the groups of largest and smallest range. These curves are almost exactly of the same type as the others, showing that the curves of largest amplitude are merely magnified versions of the curves of smaller amplitude.

Some unexpected correlations were found also between R and other quantities, as follows.

1. In north force the (positive or negative) excess E of the daily

mean value of the force, over the monthly mean value for the same calendar month, is less on quiet days of large S$_q$ range (R) than on days of small range. The average value of the north force on quiet days is greater than the average value from all days (9.5), the excess for the days considered being $2\cdot4\gamma$ at Eskdalemuir; but on dividing the days into three nearly equal groups according to range, the average value
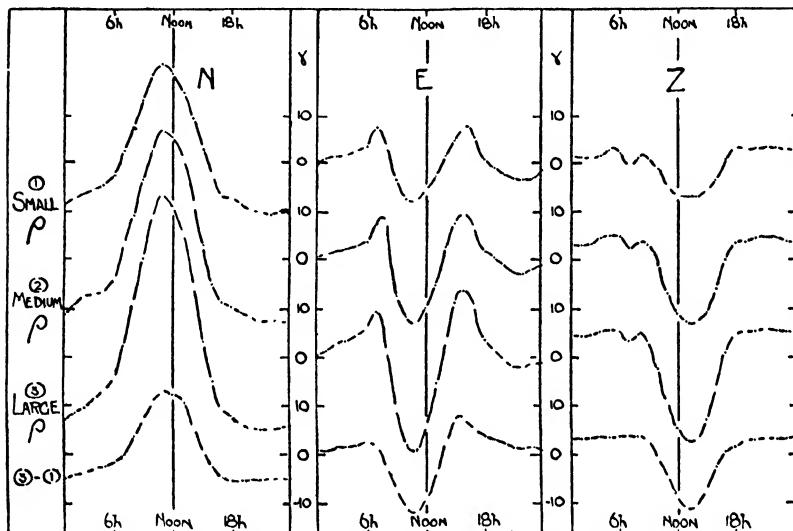
FIG. 18. The daily variations on quiet days, at Batavia, June–July 1913–23; north, east, and vertical components. Averages for days of small, medium, and large range, as in Fig. 17

of E for the small-range group was found to be $3\cdot3\gamma$, and for the large-range group only $1\cdot5\gamma$. The reality of this difference was confirmed by verifying that it was present in numerous subdivisions of the material according to season and sunspot epoch.

2. The non-cyclic variation (2.14 and 16.2), i.e. the net increase (d) in north force from the commencement to the close of a quiet day, is, on the average, positive, its mean value at Eskdalemuir being $2\cdot4\gamma$; it was found that d was least for the small-range group of days, its mean value being $1\cdot8\gamma$, while for the large-range group it was $2\cdot6\gamma$; this difference, though small, was fairly systematic. In vertical force d is negative, i.e. there is a decrease of vertical force on the average, on quiet days; for Eskdalemuir the mean value is $-0\cdot5\gamma$; but in the small-range group the mean d was $-1\cdot1\gamma$, and in the large-range group it was $+0\cdot3\gamma$, showing an algebraic increase of d with R, and consequently a reversal of sign in d.

In a later paper [20] Chapman and Stagg extended their work to a larger group of observatories, adding those of Ebro, San Fernando, Batavia, and Mauritius. They found that in almost all cases the correlation between the departures of the daily range from its normal value, in different elements and at different observatories, was a positive one, but its value was less the more distant the pair of stations. They noted also that the very quiet days often occur in sequences of two or more, and that there is a tendency for abnormalities of range to persist for two or more days.

These irregular day-to-day changes of intensity of the S_q variations imply corresponding changes in the overhead current-systems which, according to §§ 6 and 7, would account for the S_q field. If the intensity-changes were similar in all elements at all stations, the current-system would preserve a constant form, only changing in intensity. The fact that the S_q intensity-changes vary from element to element, and from station to station, implies some change of form and of relative intensity of the current-system in different places. The positive sign of the correlations above referred to indicates that in the main the current-intensity everywhere increases or decreases together, though by different amounts.

The variability of the harmonic coefficients for the solar (and lunar) daily variations, which bears closely on the variability of S_q, will be discussed in Chapter XIX.

7.9. Hasegawa's discussion of S_q variability. In a series of papers Hasegawa [21] has studied the day-to-day variability of the S_q field and its current-system. He showed that at stations fairly near the latitude of reversal of the daily variation in X, that is, near 40° latitude, the *type* of the X variation may change from day to day. Sometimes even on *consecutive* quiet days the X variation is of opposite type, as if the centres C and C' of the current-circuits of Fig. 15 shifted northwards or southwards. The type of the Y and Z variations, on the other hand, is much the same on such days: this would be expected from the discussion in § 6.

Fig. 19 illustrates this for two pairs of consecutive or adjacent quiet days, 1933 September 23, 24 and 1934 June 21, 23, for Aso, Japan (33° N., 131° E.). Hasegawa states that such extreme differences between the X variation on adjacent days are unusual; the X variation on most quiet days is of an intermediate type. He called the two extreme types E and P, the former (as on 1933 September 23) being similar to that at equatorial stations, and the P type (as on 1933 September 24) resembling that of stations on the poleward side of the current-centre.

He obtained the daily variation curves for a number of other observatories for the above days, and by interpolation constructed X daily curves for stations in longitude 120° E. at latitudes spaced at intervals
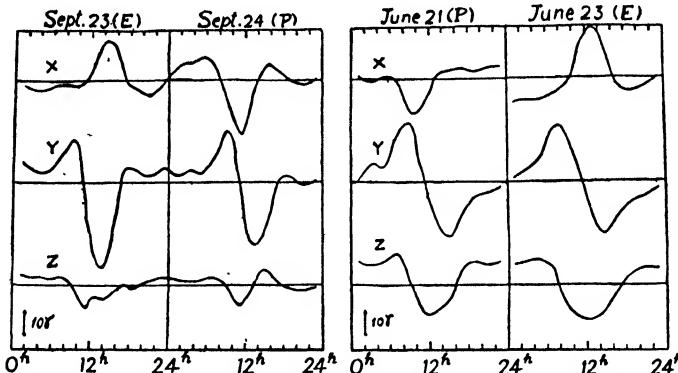
FIG. 19. Examples of two different types (P and E) of daily variation at Aso magnetic observatory (32·9° N., 131·0° E.). Two days in September 1933 (left), and two days in June 1934 (right). (After Hasegawa)
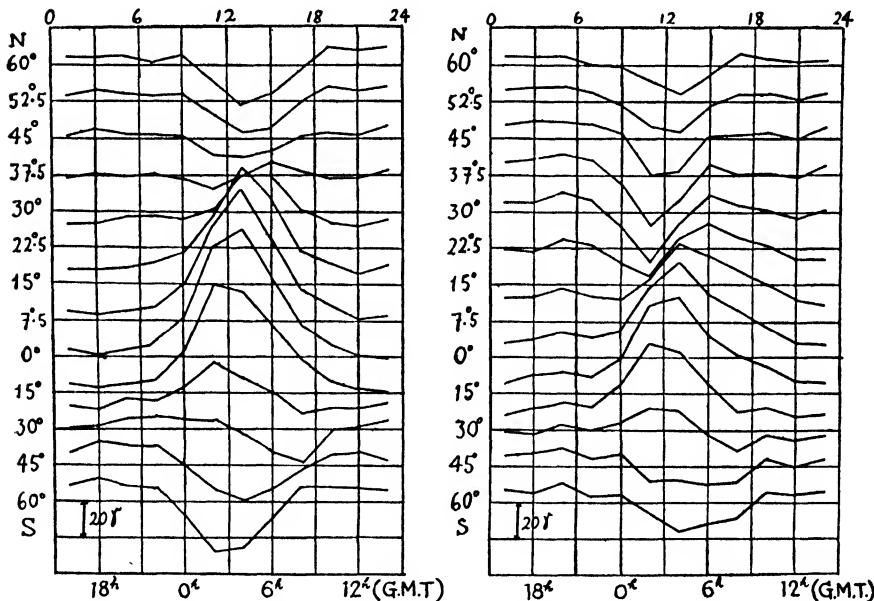
FIG. 20. Daily variations in the north component X at different points on the meridian 120° E., on September 23 (left) and 24 (right), 1933. (After Hasegawa)

from the equator of 7·5° (north) and 15° (south). These curves for 1933 September 23, 24 are shown in Fig. 20 [21 a]; the transition from the E to the P type of X variation evidently occurs at different latitudes

on the two days. On September 23 the transitional latitude (that of
the current-centre C) is at about 37° N., and on the 24th at about
22° N. On June 21 and 23 the corresponding latitudes are 27° N. and
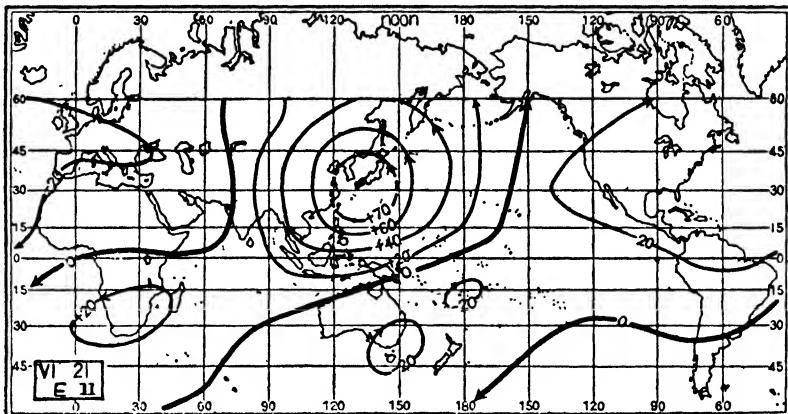
FIG. 21
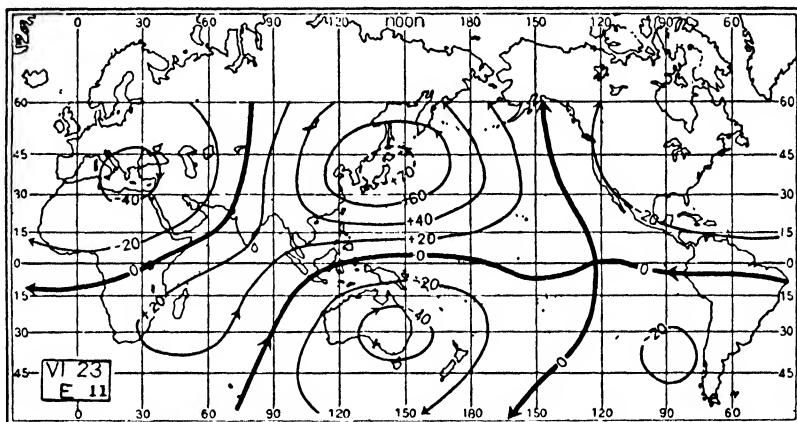
FIG. 22

FIGS. 21 and 22. The external current-system corresponding to the daily magnetic
variations, at 11ʰ local time of the meridian 135° E., or 2ʰ G.M.T., on June 21 and
23, 1934. (After Hasegawa and Ota)

43° N. These changes of latitude are very considerable—15° in the one
case, and 16° in the other.

Hasegawa [21f] considered the current-systems on the above days,
constructing them firstly by spherical harmonic analysis and afterwards
by simpler graphical methods. Figs. 21 and 22 reproduce two of his
diagrams, for 1934 June 21, 23; the lines may be regarded either as
equipotential lines of the magnetic field or as current-lines of the hypo-

thetical current-system. The numbers on the lines measure the current-flow in units of 1,000 amperes. The day-to-day change, in such cases as these, is comparable with the large seasonal changes shown by Figs. 15 and 16. The 1933 September 23 current-system was found to have more than one centre in the northern day hemisphere.

Hasegawa also made a statistical examination of various charac-
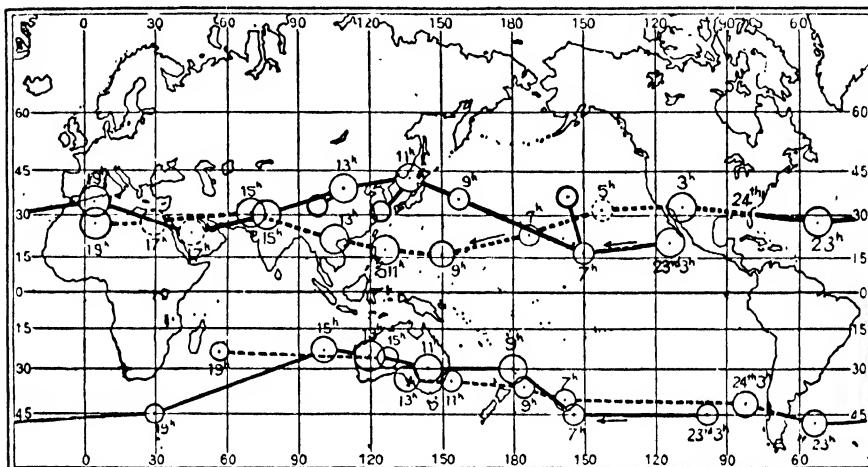
FIG. 23. Map showing the tracks of the centres of the equipotential curves of the field of the daily magnetic variation on September 23 (full line) and 24 (broken line) at various Japanese Mean Times (beginning at 3ʰ J.M.T. on September 23, i.e. 18ʰ G.M.T. on September 22). (After Hasegawa)

teristics of P and E days at Aso and Kakioka (Japan), using data for the period 1932–4. In particular, he constructed a 27-day time-pattern (p. 412) showing the temporal incidence of days of E, P, and P' (intermediate P) type over several months. This diagram is reproduced (and explained, along with other time-patterns of geomagnetic interest) on p. 412, as Fig. 12.8. It appears from this figure that either type of X variation tends to persist for some days, implying a corresponding continuance of the centres of the S_q current-circuits in a relatively high or low latitude when crossing a particular meridian, high latitude of the centre producing the E-type, low latitude producing the P-type of daily variation of X at Aso.

Hasegawa also traced the track of the current-centre round the earth over the period 1933 September 22, 18ʰ G.M.T., to September 24, 10ʰ G.M.T., showing how during its course it moved northwards or southwards: his figure is reproduced here as Fig. 23.

Hasegawa found that the occurrence of either type of X variation

appears to have no marked relationship with the magnetic activity. His papers are further discussed in 20.10.

A similar discussion of the variability of the daily variation of horizontal intensity had been previously (1932) given by Bartels for Watheroo observatory, which lies near the centre of the southern current-system; this evidence will be described in 19.4.

7.10. The large daily variation of horizontal intensity at Huancayo. The average current-systems reproduced in Figs. 15 and 16 are based on the daily magnetic variations at observatories established prior to 1920. The new observatory at Huancayo, Peru (lat. 12° 3′ S., long. 75° 20′ W.), established by the Carnegie Institution of Washington, has brought to light an important new feature which must later be incorporated in the current-diagram. This is the very large range of the daily variation S_q in the horizontal intensity H (or north component X) at Huancayo (p. 212). The noon value of H is often more than 150γ above the midnight value. The comparative magnitude of the daily variation of the horizontal components X and Y at Huancayo appears in Fig. 8. Harmonic analysis of data for a whole year (1929–30, Johnston and McNish) gives for X:

Amplitudes   . $c_1 = 53 \cdot 2\gamma$, $c_2 = 28 \cdot 6\gamma$, $c_3 = 12 \cdot 5\gamma$, $c_4 = 12 \cdot 5\gamma$.
Phases .   . $\epsilon_1 = 282°$, $\epsilon_2 = 106°$, $\epsilon_3 = 304°$, $\epsilon_4 = 163°$.

The amplitudes c_1 and c_2 are at least twice as big as those found at other tropical stations (see Table 2, p. 227). Figs. 24 and 25 show the same feature, on comparison with Fig. 13. Interpreting the daily variations at five American stations in terms of the strength of the S_q current-circuit, McNish showed that, at the equinoxes and in the western hemisphere, at 11ʰ mean time of the meridian 75° W., there is a decided asymmetry with respect to the equator, the northern circuit amounting to 75,000 amperes, the southern to 160,000 amperes, more than twice the value of Fig. 15. (See also 20.9, 8.16, and 19.7.)

A conspicuous feature of the daily variation of H at Huancayo is the jagged appearance of the curve in the day-time; even on magnetically quiet days the record shows fluctuations of several γ amplitude with periods of about a minute or more (Fig. 10, p. 266).

7.11. The zonal part of the S_q field. The S_q variation is generally reckoned from the daily mean value of each element at each station as basis, but, as Schmidt has pointed out [27], this is only a convenient and not an essential basis. Indeed there is reason to think that S_q might more properly be reckoned from the (quiet) night value of each

element.   The Huancayo daily variation curves of Figs. 24, 25, and 6.4 specially suggest that the daily change from the night value may alter in magnitude from day to day, while preserving its general character, and that the night value remains approximately constant throughout these changes (see particularly the *H* curves for the last ten days of
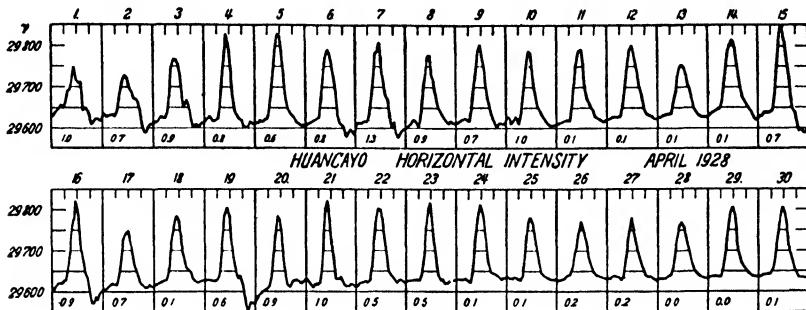
FIG. 24
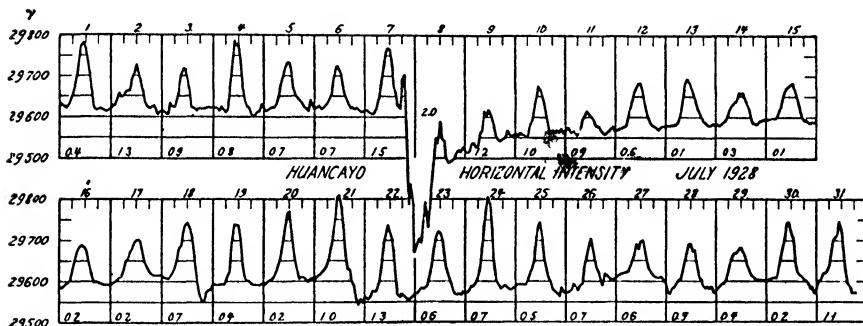
FIG. 25

FIGS. 24 and 25. Huancayo horizontal intensity. Consecutive hourly means for a quiet month (April 1928) and a disturbed month (July 1928)

April 1928).   Fig. 8.11 also strongly suggests that the S_q current-system dies away rapidly after nightfall, and that it is inappropriate from a physical standpoint to reckon the daily variation and the current-system from the daily mean value.   If the night value were adopted as the datum-line for S_q, the current-systems shown in Figs. 15 and 16 must be modified by the addition of 'zonal' currents flowing along the parallels of latitude. This point has been discussed by Hasegawa [21].

THE LUNAR DAILY MAGNETIC VARIATION (L)

8.1. The computation of L. In 1839 Kreil, of Prague, sought for a regular periodic change in the earth's magnetic field depending on the lunar hour angle [1], but it was not before 1850 (see 26.15) that he certainly detected such a variation. It is in general small compared with the solar daily variation S, a fact which, together with the nearly equal length of the solar and lunar days, renders the lunar daily variation somewhat difficult to determine.

Perhaps the simplest method of determining the lunar daily variation in any element at any station is the following; this is also the method by which most of the determinations of L hitherto published have been derived, though where Hollerith tabulating and sorting machines are available, other more rapid and accurate methods can be used. There is also an increasing tendency to use bihourly instead of hourly values.

The monthly mean hourly values of the element are taken to represent the solar daily variation S during the month, independent of the lunar daily variation; this would be quite true were S uniform throughout the month, and were there just 29 days (a complete lunation) in the month, for then in each month every lunar hour would occur equally at each solar hour in turn, and L would be averaged out. When there are more or less than 29 days in the month, L is not quite averaged out from the monthly mean hourly inequality, but the part of L remaining is very small. The monthly mean value of the element for each hour is subtracted from all the values at that hour during the month; this leaves a set of differences, one for every hour in the month, giving the variation of the element freed from the average solar daily variation. The changes in these differences from hour to hour are due to disturbance, to the secular variation, and to residual solar daily variation (owing to its changes from day to day, 7.8), as well as to the lunar daily variation.

These differences are regrouped according to lunar time, and it is assumed that if a sufficient number of days are taken the non-lunar variations will average out, even though on individual days some of them may exceed the lunar daily changes; the general manner in which these non-lunar variations average out will be discussed in 16.29.

The length of a lunar day varies slightly (5.2) about its average value, approximately $24^h 50^m$ in mean solar time: it has therefore often been

found convenient to treat it as being of duration 25 solar hours. Each lunar day at the given station is regarded as beginning at the *lower* local transit of the moon (lunar midnight). This usually does not fall at an exact solar hour, and the nearest solar hour must be used instead. The difference for this hour, and those for the 24 succeeding hours, are written down in a row on a 'lunar sheet', to represent the lunar daily variation (plus 'accidental' variations) for that lunar day; it is convenient to add also a further hourly difference at the end of the row, making 26 entries in all, for the purpose of subsequently eliminating the non-cyclic variation, the amount of which may be quite comparable with the range of L.

The solar times which are adopted as most nearly corresponding to the lunar hour 0 generally differ from one day to the next by 25^h, so that the 26th entry on the lunar row for one day is the first on that for the next lunar day; occasionally, however, the interval is only 24^h, and then the 25th and 26th entries for the earlier lunar day are the first and second for the next day. The simplest plan is to write all the rows corresponding to the lunar days falling within one calendar month on one sheet. Totals and means for the 26 columns can then be formed, and the 26 resulting hourly means, reduced to 25 by elimination of the mean non-cyclic variation (2.14), will give the mean lunar daily variation for that month. In general it will be affected by considerable accidental error, and it is necessary to combine the results for many months in order to get a well-determined lunar daily inequality.

8.2. The average L for a full lunation. When determined in this way from the mean of a number of months (or, what is nearly equivalent, from the mean of several whole lunations), the lunar daily variation is found to be always of the same simple character for every element and station: it consists of a regular semi-diurnal oscillation or repeated wave. Other harmonic components of relatively small amplitude may be present, but their lack of regularity and consistency (when reductions are made for different but similar groups of months) proves that they are accidental and no real part of the phenomenon.

8.3. The change of L with the moon's phases. It was discovered by Broun [2] (from the study of the declination at Trevandrum, 1854–64) that this simple character belongs to L only when determined from the mean of a number of lunations, and that at any particular epoch in the lunation the lunar daily variation is not merely a repeated wave. He determined L from the mean of a large number of days all occurring at nearly the same phase in different lunations—for example, lunar

days within one or two days of new moon (and similarly for seven other groups centred at one-eighth phase, first quarter, three-eighths phase, full moon, and so on). On plotting the corresponding hourly irregularities he found that the curves indicated a regular cycle of change in L during the lunation. Each curve showed approximately two waves, but of unequal amplitude, and the part of the curve which showed the greatest movement occurred earlier and earlier in the lunar day as the lunation progressed. Now the hours of daylight, occupying about half of each lunar day, regress in this way during the lunation; for example, at new moon, lunar hour 0^h approximately coincides with solar hour 0, and the centre of the daylight period is at roughly 12^h lunar time, while at first quarter the lunar hour 0 corresponds to solar hour 6, and the hours of daylight are centred at about 6^h lunar time. On redrawing the eight curves so that the abscissae referred to solar time, it appeared that the chief movements occurred during the hours of daylight. Similar results were afterwards obtained independently by Chambers [3], from the study of D and H at Bombay, 1846–72. Fig. 1 gives results for Batavia, after Chapman [9].

FIG. 1. The lunar daily variation of the magnetic west declination at Batavia: for eight phases of the moon, beginning with new moon, and for the mean of all phases (lowest curve). The variations during the daylight hours are shown by thick lines, those during the night by thin lines. Scale: The vertical distances between the zero lines of the eight upper curves correspond to 4γ in west force, or about $0 \cdot 37'$ in direction

Broun [2] proceeded to form two mean lunar daily inequalities from a number of lunations, combining lunar days from all phases, but making separate sums and means from those hourly values which corresponded to the night and the day respectively. The two inequalities both proved to be purely semi-diurnal, but in amplitude the day curve was five times as great as the night curve.

When the curves representing the inequalities for the separate phases of the moon are arranged according to solar time, the large movements

occur during the day hours in each case, but they are not in the same sense at the same solar hour at different lunar phases; when averaged throughout the lunation, according to solar time, they vanish. This, indeed, must be so, because the average solar daily variation has already been abstracted before retabulating the hourly differences.

8.4. Moos's expression for the change of L in the course of a lunation. When the lunar daily curves or inequalities are arranged according 'to lunar time, the average, as has been seen, is a repeated wave. The forms of the curves for the separate phases suggest that they might be derived from this mean curve by magnifying that part of it which falls during the hours of daylight (Fig. 1). Moos [4] expressed this idea mathematically; taking the mean curve as represented by the expression

$$c \cos(2\tau + \epsilon), \tag{1}$$

where τ denotes lunar time measured in angle at the rate 2π per lunar day, he proposed to replace the constant amplitude-factor c by a factor c', which itself varies in the course of the lunar day, in such a way that c' is greatest, at each phase, during the hours of daylight. He did this by writing

$$c' = c(1 - a \cos t), \tag{2}$$

where t denotes solar time measured in angle at the rate of 2π per solar day, from midnight. Thus at midday ($t = 12^h = \pi$), $c' = c(1+a)$, and at midnight ($t = 0^h = 0$), $c' = c(1-a)$. Moos for simplicity took a to be unity, and drew curves showing the variation of $c' \cos(2\tau + \epsilon)$, that is, of

$$c(1 - \cos t)\cos(2\tau + \epsilon), \tag{3}$$

where c and ϵ were determined from the mean curve; he compared these curves for the separate phases with the observed curves, and found that they were of very similar form, thus tending to confirm his idea that the lunar daily variation is produced by the moon but that its amplitude is largely determined by the sun.

8.5. Chapman's expression for L; the phase-law. Since the earth performs one more revolution relative to the sun than relative to the moon during a lunation, t increases by 2π more than τ during this period. If we write (p. 163)

$$t = \tau + \nu, \tag{4}$$

ν increases from 0 at one new moon to 2π at the next, and forms a convenient measure of the epoch in the lunation (being, for example, $\frac{1}{4}\pi$, $\frac{1}{2}\pi$, π,... at one-eighth phase, first quarter, full moon, and so on). Hence the formula of Moos can be written (in a manner corresponding more closely to that actually given by him) as

$$c\{1 - a \cos(\tau + \nu)\}\cos(2\tau + \epsilon). \tag{5}$$

Chapman [5] made the simple transformation of this to the form

$$\tfrac{1}{2}c\{-a\cos(\tau+\epsilon-\nu)+2\cos(2\tau+\epsilon)-a\cos(3\tau+\epsilon+\nu)\}, \qquad (6)$$

and saw that this implied certain features which, if the formula of Moos were true, should be shown by the Fourier constants of L at each phase of the lunation: namely, that there should be three harmonic components, the semi-diurnal one being of constant phase-angle, while the amplitudes of the first and third components should be equal, and their phase-angles should be $\pi+\epsilon-\nu$ and $\pi+\epsilon+\nu$. Since ν changes (increases) by 2π during a lunation, the phases of these first and third components should respectively decrease and increase by 2π per lunation. He analysed the inequalities given by Broun, Chambers, and Moos, and others given by Figee [10] for Batavia; it appeared that (subject to some accidental error) the *amplitudes* of the various harmonics were constant throughout the lunation, in the average of a number of lunations: the phase-angle also was constant for the second component, as was expected; and the phase-angles of the first and third components showed the expected decrease and increase, respectively, by 2π, in the course of the lunation. In these respects the formula by Moos was confirmed, but in others not so: the phases of the first and third components at new moon ($\nu = 0$) were not equal to one another, and 180° different from that of the second component; nor were the first and third amplitudes equal. Moreover, a fourth component was present (and doubtless others), and Chapman found that its phase increased by 4π per lunation, though owing to the small amplitude of this component the law of variation of its phase was often obscured by accidental error. This component and its phase-change, though discovered empirically, can be deduced from a formula similar to that of Moos by including a term $bc\cos 2t$ or $bc\cos 2(\tau+\nu)$ in c'; this introduces two additional terms, $\tfrac{1}{2}bc\{\cos(\epsilon-2\nu)+\cos(4\tau+2\nu+\epsilon)\}$, into $c'\cos(2\tau+\epsilon)$. But though Moos's formula is suggestive, it is too simple, even when thus modified, to fit the facts concerning the relative amplitudes and phase-angles of L (at new moon, for example). The expression for L in any element, as found by Chapman (up to the fourth harmonic), cannot be written less generally than

$$\sum_{\sigma=1}^{4} c_\sigma \sin\{\sigma\tau+\epsilon_\sigma+(\sigma-2)\nu\}; \qquad (7)$$

that is, no general relations between the c_σ's and the ϵ_σ's, for different values of σ, have been found. Owing to the steady change of phase by 2π or 4π during the lunation, all components save the second, which preserves its phase constant, average out in the course of a lunation.

Since $t = \tau + \nu$, the above expression may also be written

$$\sum_\sigma c_\sigma \sin(\sigma t + \epsilon_\sigma - 2\nu), \tag{8}$$

indicating that, if the inequality is expressed (at any lunar phase) in terms of the *solar* time t, all the components regress in phase-angle by the same amount, 4π, per lunation, so that upper or lower transit (for which ν is different by 2π) can be used indifferently.

Finally, it should be mentioned that in many older papers the *upper* transit of the moon has been taken as beginning the lunar day; this leaves ϵ_2 and ϵ_4 unchanged, but changes ϵ_1 and ϵ_3 by 180°. In this book *lower* transit has been taken as the epoch $\tau = 0$.

8.6. Chambers's luni-solar variation. Chambers [3] showed that at any lunar phase (ν) L could be expressed relative to the solar time t in the form

$$f_{c,2}(t)\cos 2\nu + f_{s,2}(t)\sin 2\nu, \tag{9}$$

where $f_{c,2}(t)$ and $f_{s,2}(t)$ denote the lunar daily variation, referred to solar time, at new moon and one-eighth phase respectively. He verified this formula for the determinations of L for Bombay (D and H), Trevandrum (D), and Batavia (D). This simple expression for L, in all its changes throughout the lunation, led him to conclude that L is really a solar daily variation, and not a lunar daily variation (presumably because it admits of this simple representation in terms of the solar time t); but the supposed solar daily variation undergoes cycles of wave-like change in the lunation, and therefore depends on the relative positions of the sun and moon; he therefore called it a *luni-solar variation*. Though the name itself has some justification, the conclusion that L is primarily a solar daily variation is difficult to reconcile with the presence in it of one harmonic component (the second) which, referred to lunar time, has a constant phase throughout the lunation, whereas there is no harmonic of constant phase relative to solar time.

Chambers's functions $f_{c,2}(t)$ and $f_{s,2}(t)$ are given, according to (8) (putting $\nu = 0$ for new moon to get $f_{c,2}$, and $\nu = \frac{1}{4}\pi$ for first-eighth phase to get $f_{s,2}$), by

$$f_{c,2} = \sum_\sigma c_\sigma \sin(\sigma t + \epsilon_\sigma), \tag{10}$$

$$f_{s,2} = \sum_\sigma c_\sigma \sin(\sigma t + \epsilon_\sigma - \tfrac{1}{2}\pi)$$

$$= -\sum_\sigma c_\sigma \cos(\sigma t + \epsilon_\sigma). \tag{11}$$

On substituting these in Chambers's formula we get

$$\sum_\sigma c_\sigma\{\sin(\sigma t + \epsilon_\sigma)\cos 2\nu - \cos(\sigma t + \epsilon_\sigma)\sin 2\nu\} = \sum_\sigma c_\sigma \sin(\sigma t + \epsilon_\sigma - 2\nu), \tag{12}$$

which agrees with (8). But the fact that Chambers did not discover

the simple relation between $f_{c,2}$ and $f_{s,2}$, corresponding to the above equations (10) and (11) (i.e. that their harmonic components have equal amplitudes and differ in phase by $\frac{1}{2}\pi$), made his able investigation unfruitful; the further advances began with Moos's simpler and more direct formula for L. It may also be mentioned that Chambers went on to add more terms to his formula (in the hope of further improving its agreement with observation) which were such as implied the existence of a daily term $\sin(\tau + \epsilon')$, of constant phase during the lunation, in L; but there seems to be no evidence for the existence of this term.

8.7. The calculation of L for different transit times. Owing to the continual change of phase of all the components of L save the second, it is convenient, in retabulating the hourly differences (§1) according to lunar time, to write the lunar daily rows on different sheets according to their solar hour of commencement, as the latter affords a simple measure of the lunar age ν. All the days at a given age of the moon, so specified, can in this way be conveniently grouped together, to give the mean lunar hourly inequality at that epoch of the lunation.

8.8. L in Greenwich declination. The most reliable determination of L yet obtained for any element and station is that for D at Greenwich, based on sixty-three years' data. The twenty-four inequalities for L at different lunar ages ($\nu = 0$, 1, 2,..., 24 lunar hours), derived from the whole of this material, were each harmonically analysed in order to test the law of change of phase for the different components. The results are indicated in Fig. 2, which illustrates the values of ϵ_σ derived from the twenty-four inequalities. If the above-mentioned phase-changes occurred exactly, without being affected by accidental errors, the phases should change at the rates indicated by the straight lines (the scale for ϵ_4 is half that for ϵ_1 to ϵ_3, in order that the phase-line may be shown with the same slope as for ϵ_1 and ϵ_3; the reduction is also appropriate because of the very small amplitude c_4 ($= 0 \cdot 1 \gamma$ in the mean), which makes the significance of a given accidental variation of ϵ_4, in force units, much less than that of the same variation in ϵ_1 to ϵ_3). It is clear from the figure that, in spite of many accidental irregularities, the phase-angles do change at the rates and in the directions mentioned [9]. Another confirmation of the phase-law will be given in 19.7.

8.9. Horizontal vector diagrams for L. The above phase-changes are found in all three elements at all stations for which L has been determined. The fact that they correspond to an accentuation of the lunar daily changes during the hours of sunlight is particularly well

brought out by the vector diagrams (7.2) at each phase. In the mean of a lunation, since both in D and H or in X and Y the variation is purely semi-diurnal, the vector diagram consists of an ellipse twice
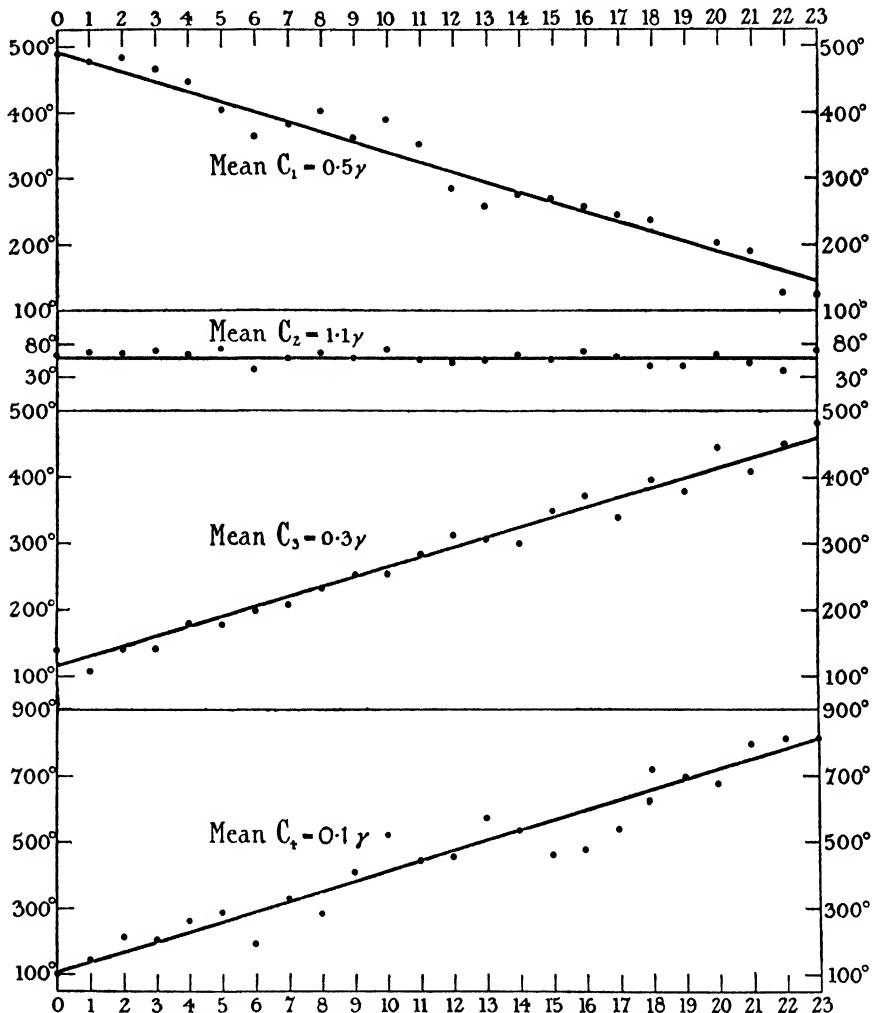
FIG. 2. The monthly changes in the phase-angles ϵ_1, ϵ_2, ϵ_3, ϵ_4 of the sine-terms $c_1\sin(\tau+\epsilon_1)$, $c_2\sin(2\tau+\epsilon_2)$, etc., in the Fourier expression for the lunar daily variation of Greenwich west declination (mean for 63 years); the abscissa ν is 0 at new moon, and 24 hours at the next new moon

described in the lunar day. But at each particular epoch in the lunation the part of the curve corresponding to the hours of daylight is magnified and the remainder is reduced. This is shown in Fig. 3, which gives the vector diagram for Pavlovsk (summer) for the mean of the lunation

and for eight lunar phases, the daylight part of the latter curves being drawn in thicker lines than the night part.

The diagrams for the eight phases show that the magnification of the lunar changes begins at sunrise and affects all parts of the lunar
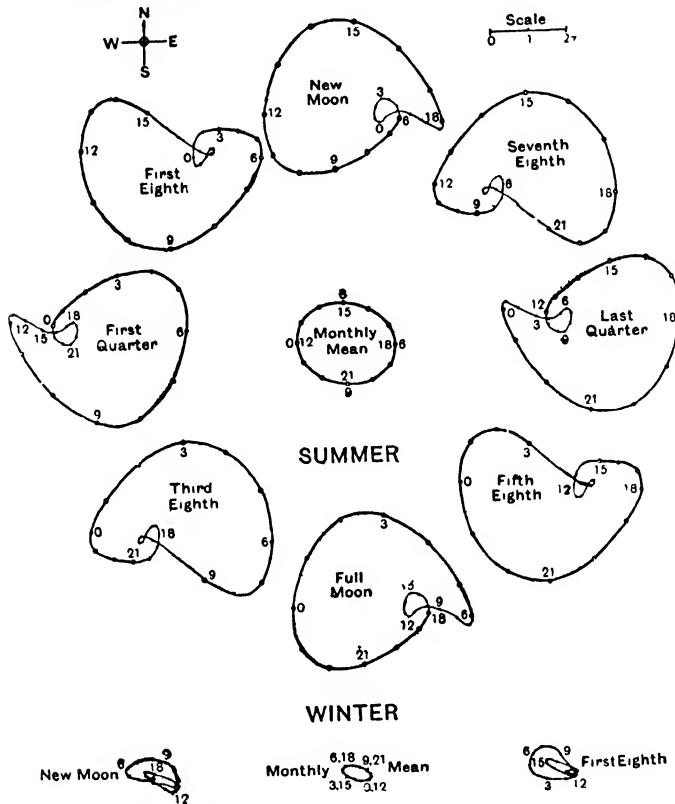
FIG. 3. Vector diagrams for the lunar daily variation of the horizontal magnetic force components at Pavlovsk, in summer (above) and winter (below): for the mean of a number of whole lunations (in the centre), and also at various particular epochs in the lunation, the daylight part of the latter curves being drawn in thicker lines. Scale for 2γ in upper right corner

day in turn at different lunar ages, but that, though the sun has thus an important influence on the intensity of the changes, their direction is chiefly governed by the moon. This is evident on comparing the direction of description of the vector diagrams at any lunar hour of daylight and at any lunar age, with the direction at the same hour in the monthly mean diagram.

The eight diagrams round the centre of Fig. 3 evidently consist of only two distinct types, differently orientated at different lunar ages.

Hence, in adding similar diagrams for the winter, only two diagrams for particular lunar ages are given.

8.10. Subdivisions of the data for L. The winter and summer values of L are, of course, obtained by separating the lunar days not only into twenty-four groups according to the solar hour of commencement, but also into further subgroups. Such subdivisions are conveniently made for each individual month, though the results may afterwards be combined into seasonal results. Other subdivisions which have been made by various writers relate to the distance of the moon, its declination, the magnetic activity prevailing during each lunar day, and the sunspottedness of the year in which the day falls.

8.11. The influence upon L of changes in lunar distance. The influence upon L of changes in lunar distance has been investigated by Broun [2] (Trevandrum), Figee [10] (Batavia), Schmidt [11] (Potsdam), and Chapman [8, 9]. Broun and Figee, who considered only the second harmonic component (the semi-diurnal component, of constant phase), both found that its amplitude is increased as the lunar distance diminishes, but whereas Broun concluded that the increase was in the ratio of the moon's tide-producing force, that is, inversely as the cube of the moon's distance d, Figee found a larger ratio. The ratio $(1/d)^3$ is 1·39 as between the actual days of perigee and apogee, or 1·23 when half-lunations centred about these epochs are considered. Owing to the small amplitude of L, the material used by Broun and Figee was inadequate to establish the exact value of the changes due to varying lunar distance, and even the much larger mass of data used by Chapman does not suffice to determine it more accurately than with an error of about 5 per cent. of the whole amplitude. His weighted mean results, obtained from the investigation of the three elements at many stations (and depending in the aggregate on about eighty years' hourly observations), were as follows, for groups of five days centred at apogee (A), five centred at perigee (P), with the remaining days (a, p) in each of the half-lunations centred at these epochs:

<div align="center">

TABLE 1

</div>

Group	No. of days used	Percentage deviation of c_2 from the mean	Phase-change from the mean
P	4,898	+15	+9·5°
p	8,279	┘ 9	−0·3°
a	8,307	−7	−3·7°
A	4,923	−18	−6·0°

The change between apogee and perigee is thus found to be 33 per cent., whereas the change of the tide-producing forces from the P to the A groups of days is in the ratio 37 per cent. By considering the half-lunations, for which the mean ratio $(1/d)^3$ is 1·23 (corresponding to a change of 23 per cent.), changes of about 21 per cent. in c_2 were found. In obtaining the above figures, regard was had only to elements, seasons, and stations for which c_2 (the Fourier coefficient of the second component of L) was at least 1γ.

Chapman found also that the increase of semi-amplitude c_σ from apogee to perigee was shown also by the other harmonic components, though as these are usually of smaller magnitude than c_2 it is not yet possible to give a reliable determination of the amount of the percentage change for them.

In addition to the change of amplitude, all the components show a change of phase-angle from apogee to perigee; for the second component this is indicated in the above table. It amounts to about 16°, ϵ_2 being greater at perigee than at apogee; when half-lunations are considered the change is about 10°. The other components show similar changes, but in all the components even the best determinations so far made are liable to a good deal of accidental error. Hence the reality of the phase-change is perhaps still doubtful and needs further examination.

In fact L depends so much on so many different factors that it is difficult to disentangle its various changes, particularly in view of the small magnitude of the whole; but it seems possible to say that its amplitude varies approximately in the ratio of the moon's tide-producing force, that is, as $(1/d)^3$; so far as the present results go, however, the index 3 might be replaced by a number slightly less than 3, though certainly not by 2. It is not yet possible to say whether the dependence of L on the lunar distance is different for different elements, stations, or seasons, but the results hitherto obtained are consistent with the view that it is not, since the differences found are not more than can reasonably be ascribed to accidental error.

8.12. The influence of the magnetic activity upon L. Another factor on which L depends is the magnetic activity. This appears to have been first noticed by Figee [10], who found that at Batavia the amplitude of L was greater when the hourly values at disturbed hours were included in the reductions than when they were excluded (as was the custom of Sabine [14] and his many followers). Van Bemmelen [12] investigated the matter further, and found that the inclusion of the

disturbed hours leads to values of the amplitude c_2, for L, about 1·5 times as great as would otherwise be found, without appreciably affecting the phase ϵ_2. The ratio of the amplitudes would be expected to depend on the degree of disturbance chosen as the limit beyond which an hour was considered disturbed. The question has been examined in considerable detail by Chapman, who classified the lunar days into four groups according to the degree of disturbance present, as measured by the disturbance daily range (R_D, p. 357) in the series of hourly

TABLE 2. *Amplitudes of lunar sine-waves (unit* $0·01\,\gamma$) *for various degrees of magnetic activity*

		Greenwich declination								
		Yearly mean					Summer			
Group	Mean R_D	No. of days	c_1	c_2	c_3	c_4	No. of days	c_1	c_2	c_3
XX (very disturbed)	85 γ	736	109	164	64	17	..	..	..	..
X (disturbed)	48 γ	3,575	39	143	29	22	1,123	106	233	50
O (average)	23 γ	12,176	53	176	40	9	4,069	118	187	64
Q (quiet)	11 γ	4,275	34	75	23	4	1,627	62	106	26

	Equinox					Winter			
Group	No. of days	c_1	c_2	c_3	c_4	No. of days	c_1	c_2	c_3
XX (very disturbed)									
X (disturbed)	1,439	44	120	53	35	1,013	79	98	21
O (average)	3,595	68	114	56	21	4,512	34	73	6
Q (quiet)	1,559	28	71	37	14	1,089	8	56	1

	Batavia declination (Nov.–Feb.)					Zikawei declination (May–Aug.)				
Group	No. of days	c_1	c_2	c_3	c_4	No. of days	c_1	c_2	c_3	c_4
Disturbed (X)	641	388	542	368	111	581	263	445	229	37
Moderately d. (O_2)	966	190	317	234	74	1,246	156	308	174	31
Moderately q. (O_1)	1,123	75	164	124	41	1,303	54	162	34	3
Quiet (Q)	431	18	52	56	22	431	2	69	26	9

	Batavia H (Nov.–Feb.)				Pavlovsk H (May–Aug.)				Pavlovsk Z (May–Aug.)			
Group	No. of days	c_1	c_2	c_3	No. of days	c_1	c_2	c_3	No. of days	c_1	c_2	c_3
Disturbed (X)	478	169	163	85	669	65	147	97	741	155	49	33
Moderately d. (O_2)	537	138	180	73	1,063	102	96	84	565	49	21	4
Moderately q. (O_1)	1,051	61	120	45	1,095	111	118	74	1,603	34	13	10
Quiet (Q)	918	50	78	33	1,533	55	29	52	1,231	28	3	3

differences for each lunar day, after the monthly mean solar daily variation had been subtracted. Usually four groups of days of different degrees of disturbance were considered. The semi-amplitudes c_σ of the various harmonic components were found to increase with the intensity of magnetic disturbance, in ratios which vary considerably from one element, station, season, and component to another; in some cases there was a suggestion of systematic changes of phase-angle.

In considering the results in Table 2, the relative numbers of days in different subdivisions of the same data will serve to indicate the relative range of disturbance in each group. As regards the values of c_σ, in all cases they are expressed in the unit $0 \cdot 01\gamma$ ($10^{-7}\Gamma$). The mean disturbance ranges R_D (in γ) for the four groups are also indicated.

The dependence of c_σ on the degree of disturbance is shown in Table 2 most notably in the declination at Batavia and, to a less extent, at Zikawei, the two stations (among those considered) nearest the equator; ratios as high as 10 or even more are found, between the disturbed- and quiet-day values of c_σ. At Greenwich and Pavlovsk the change in c_σ for D is less, but still striking, the extreme ratio found being about 2; H both at Batavia and Pavlovsk showed a similar ratio. In Pavlovsk Z the coefficients c_σ are too small for much reliance to be placed upon their ratios. From the Greenwich results it would seem that c_σ does not vary linearly with R_D, the rate of increase of c_σ with R_D diminishing as R_D increases.

The phases for the different groups of days are given in the following table for a few of the above groups, namely, those for which more than one value of c_σ is at least 1γ.

TABLE 3. *The phase-angles ϵ_n in the lunar daily variation*
$$\sum c_n \sin(n\tau + \dot{\epsilon}_n)$$
at new moon; $\tau =$ lunar time from lower transit

	West declination										Horizontal intensity	
	Greenwich			Batavia			Zikawei			Batavia		Pavlovsk
	Summer	Equin.	Year	Year			Year			Year		Year
Group	ϵ_1	ϵ_2	ϵ_2	ϵ_2	ϵ_1	ϵ_2	ϵ_3	ϵ_1	ϵ_2	ϵ_3	ϵ_1	ϵ_2	ϵ_2
Disturbed X	..	..	..	59°	75°	275°	123°	253°	90°	281°	349°	199°	5°
O$_2$	286°	69°	79°	65°	60°	267°	112°	248°	88°	285°	359°	203°	10°
O$_1$	287°	71°	63°	58°	38°	260°	109°	..	84°	..	..	206°	8°
Quiet Q	294°	68°	57°	55°	..	..	..	..	..	..	..	..	..

The best determined values of ϵ are probably those for D at Batavia and Zikawei, owing to the magnitude of their coefficients c_σ; from these it would seem that ϵ increases slightly as the disturbance increases. This is confirmed by the appearance of these phase-changes in subgroups of the same material, taking years of high and low sunspottedness separately [9].

All these results are subject to some uncertainty, however, because of the fact that the solar daily variation was subtracted from the data at the outset, in the form of the monthly mean solar daily variation S. When the days are grouped according to the magnetic activity, however, the mean solar daily variation for the group will in general differ from the monthly mean S, because S for the different groups will include different amounts of S_D. Hence S is not fully eliminated, and the residual amount remaining in the lunar hourly differences is systematically different for the different activity-groups. If a very large amount of material were available, the rearrangement of the hourly differences according to lunar time, and the subsequent summation, should cause the residual S_D to be averaged out; but there is a possibility that this was not achieved in some, at least, of the above reductions. To avoid this difficulty, it is now considered best, in determining L, not to remove S from the data at the outset, but rather to determine and remove S as it appears in each subdivision of the data. The danger of the incomplete removal of S by averaging out, when the material used is not really extensive, is well shown by Chapman's published results [7] for different activity-groups of days for Pavlovsk D; in this determination S was, by oversight, not removed at all from the data, and the results for L indicated a reversal of phase from days of low to those of high activity. In a later (unpublished) reduction the appropriate S for each activity-group of data was removed in the final stages; the Pavlovsk D results, thus corrected, no longer showed any reversal of phase from low to high activity.

8.13. The influence upon L of the sunspot number. Soon after Sabine and Lamont had discovered the large change in S from year to year, which follows the variation of the annual mean sunspot numbers s, a similar effect was sought for in L. Different investigators reached opposite conclusions on the matter: Kreil, Broun, and van der Stok decided in favour of a sunspot-cycle variation in the amplitude of L; Sabine, Chambers, and Figee formed the contrary opinion. In 1914 Chapman [21] reviewed the relevant available evidence from nearly all the determinations of L hitherto made, and again discussed it in 1917 [22]

with the aid of further data obtained by himself; the variations of L from year to year at different periods and for different elements and stations were not very accordant; whereas L sometimes seemed to vary in parallel with s (though to a slighter degree than S), in other cases it varied irregularly or even inversely with s. The combined evidence was, in fact, insufficient to lead to a definite conclusion, but it seemed likely that L varied with s, to a slight extent only. This conclusion is confirmed by more recent evidence.

In the determination of L from sixty-three years' observations of D at Greenwich the data were subdivided into five groups of years classified according to sunspottedness. The following are the values of c_1, c_2, and c_3 for L, and also for S, for these five groups: the table is of interest also in indicating the relative magnitude of these components in the solar and lunar daily variations.

TABLE 4

Group	Mean sunspot number	No. of days used for L	Greenwich declination (63 years) (unit 1γ)							
			c_1		c_2		c_3		$c_1+c_2+c_3$	
			L	S	L	S	L	S	L	S
α	96	3,705	0·63	19·4	1·21	12·5	0·30	4·8	2·14	36·7
β	63	3,665	0·51	15·5	1·28	9·9	0·36	4·6	2·15	30·0
γ	44	4,623	0·48	13·6	1·11	9·3	0·36	3·7	1·95	26·6
δ	19	4,047	0·49	12·5	1·07	8·4	0·22	3·7	1·78	24·6
ϵ	6	4,722	0·38	10·5	1·05	7·8	0·33	3·5	1·76	21·8

The most important component of S, namely c_1, increases by about 85 per cent. in the above table, from $s_{\min}$ to $s_{\max}$, and c_2 and c_3 increase by about 60 and 37 per cent.; the change in $c_1+c_2+c_3$ is 68 per cent. In L the changes of the three components, even of the largest, c_2, are clearly affected by accidental error, though they definitely indicate an increase at sunspot-maximum. The increase, measured by the change in $c_1+c_2+c_3$, is 22 per cent., though the estimate of the increase might vary from 15 to 28 per cent. if we assume, what is not unlikely, that any of the values of $c_1+c_2+c_3$ may be in error by $\pm0·05\gamma$.

The smallness of the increase is the more remarkable, if it is true that L is greater on disturbed than on quiet days, since disturbed days are more frequent in years of high sunspottedness. It is of interest to compare the above values of c_2 for L with values calculated on the assumption that for any given degree of disturbance c_2 is independent of the sunspot epoch, account being taken only of the change in the relative frequency of days of different magnetic activity in the different

years. The percentage of days in the activity groups XX, X, O, Q (see Table 2) in the different groups of years are as follows:

TABLE 5

	α (max.)	β	γ	δ	ε (min.)
XX	6·0	4·8	4·4	2·2	1·2
X	29·6	21·0	17·8	13·7	7·1
O	58·9	61·6	59·3	58·7	51·1
Q	5·5	12·6	18·5	25·4	40·6

Assigning weights proportional to these numbers to the values of c_2 given in Table 2 for the groups XX, X, O, Q, the following weighted means are obtained for the five groups of years:

c_2 (calc.)

α	β	γ	δ	ε
$1·24\gamma$	$1·19\gamma$	$1·15\gamma$	$1·10\gamma$	$1·01\gamma$

These agree with those in Table 4, within the probable limits of accidental error, and support the view that the magnitude of L corresponding to any given degree of magnetic activity depends but little, if at all, on the sunspot epoch. To test this the Q and X groups of days were each divided up into two portions corresponding to years of greater and less sunspottedness than the mean, and c_2 was separately determined from the four groups of days thus formed. The results were as follows:

	No. of days	Q	No. of days	X
Sunspot-maximum	1,375	$0·70\gamma$	2,681	$1·50\gamma$
Sunspot-minimum	2,900	$0·78\gamma$	894	$1·27\gamma$

It is doubtful whether the differences here found are real, especially for the disturbed days, since the determination of c_2 from only 894 days of disturbance is naturally more liable to error than the other determinations.

The question was further examined by considering a large amount of data for D and H at Batavia and Zikawei for the seasons at which L is greatest. The amplitude is particularly large in D at Batavia. The detailed results will not be given here, but the following Table 6 summarizes the results from the whole set of data, which show a fair degree of mutual accordance. It gives the sums of corresponding values

of c_1, c_2, and c_3 for Batavia declination and horizontal force, and for Zikawei declination (for instance, the first figure for c_1 is obtained as $3 \cdot 79 + 1 \cdot 67 + 2 \cdot 35 = 7 \cdot 81\gamma$).

TABLE 6

Group		No. of days	c_1 (sum)	c_2 (sum)	c_3 (sum)
			γ	γ	γ
X	s_{max}	1,250	7·81	11·34	6·65
	s_{min}	450	9·43	12·01	7·32
O_2	s_{max}	1,663	4·86	7·79	4·49
	s_{min}	1,086	4·97	8·42	5·25
O_1	s_{max}	1,866	1·84	4·37	2·45
	s_{min}	1,611	2·00	4·55	2·68
Q	s_{max}	783	1·05	2·12	1·53
	s_{min}	997	0·69	1·93	0·87

This table suggests that c_σ for L may be slightly *less* at sunspot-maximum than at sunspot-minimum for days of equal disturbance-range R_D, except possibly on the quiet days. But the conclusion can hardly be considered established as yet, and it is desirable to test it by further evidence (cf. remarks at the end of § 12). It seems probable, however, that on (equally) disturbed days the magnitude of L is not *greater* at sunspot-maximum than at sunspot-minimum.

As regards the phase-angles, these show no appreciable systematic dependence on sunspot epoch. For Greenwich the deviations of ϵ_2 from the mean are as follows for the five groups of years:

α (max.)	β	γ	δ	ϵ
$+2°$	$+5°$	$-1°$	$+1°$	$-4°$

For other stations and elements the following are the differences shown by ϵ (sunspot-maximum value less sunspot-minimum value); values are given only when the corresponding c is at least 1γ.

TABLE 7

	Batavia D			Zikawei D			Batavia H	
Group	ϵ_1	ϵ_2	ϵ_3	ϵ_1	ϵ_2	ϵ_3	ϵ_1	ϵ_2
X	$+1°$	$-5°$	$-11°$	$-5°$	$-11°$	$-19°$	$+15°$	$-16°$
O_2	$-1°$	$-4°$	$+1°$	$+3°$	$+3°$	$-1°$	$-36°$	$-12°$
O_1	..	$+1°$	$+6°$	..	$-4°$	..	..	$+5°$

The differences here shown are probably to be ascribed to accidental error.

8.14. The seasonal change of L. So far the discussion has related to the dependence of L on the lunar age, the lunar distance, the magnetic activity, and the sunspot epoch; L also depends in a marked way on the season, much more so, in general, than S. The seasonal variation can be illustrated most conveniently by means of vector diagrams. In Fig. 3 are shown the monthly mean vector diagrams for the horizontal components of L for Pavlovsk, in summer and winter; these may be compared with the diagrams for S_q for Greenwich for summer and winter in Figs. 7.6–8. It is clear from these figures that the seasonal variation of L is much greater than that of S_q.

8.15. The ionospheric current-systems for L. Another striking method of representing the seasonal variation is shown in Figs. 4–7. These are world maps of the systems of horizontal electric currents, flowing in the upper atmosphere (the ionosphere), that could give rise to the L field; they are comparable with those given in Figs. 7.15 and 7.16 for the S_q field. They may be regarded merely as giving a world-view of the relative intensities of the two fields, at different seasons and latitudes, without implying any theoretical views as to the origin of the fields; for the method of construction of the diagrams see 17.18, 20.7, and 7.7. In each case diagrams are given for the equinox and for the solstice (northern summer). They show that in their larger features the two fields have much in common (for instance, in Figs. 6 and 7 the greater intensity of the currents over the sunlit than over the dark hemisphere), though they differ in intensity and in phase. The current-system for L is weaker than that for S_q in the ratio of about 1 to 12, the total current circulating in the main circuit of L (at the equinoxes) being about 5,300 amperes, as against 62,000 amperes for S_q. Figs. 5 and 7 indicate how much greater is the seasonal change in L than in S_q, the L summer sunlit circuit (in the northern hemisphere in June) carrying a total current of 11,000 amperes, which is more than twice the strength of the equinoctial circuit.

8.16. L in Huancayo H. In 7.10 the large amplitude of S in H at Huancayo, Peru, was described; in H at this station, however, L is even more abnormal than S, that is to say, L is unusually large, and the ratio S/L is less than elsewhere, especially at the December solstice (southern summer). This is shown [19] by the table on p. 264, due to Bartels, which gives the harmonic coefficients c_n, here denoted by s_n for S and by l_n for L, for quiet and ordinary days at Huancayo, during thirteen years. The days used are those for which the international character-figure C was less than 1·2; this excludes about one day in six.
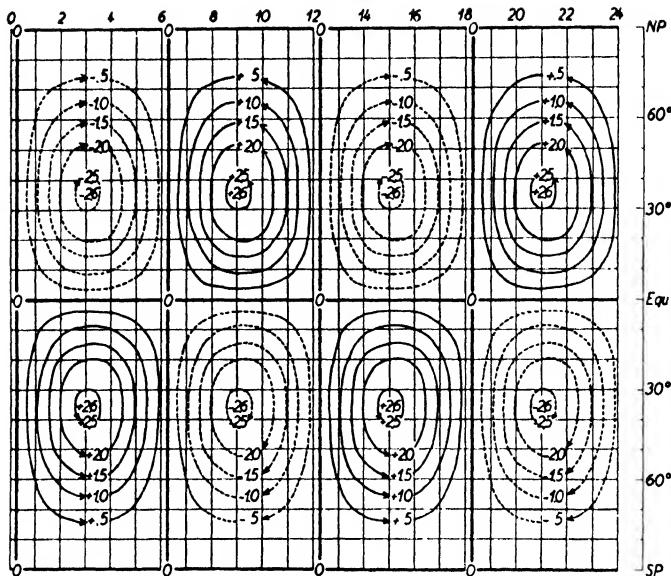
FIG. 4. Mean for a lunation, equinoxes

FIGS. 4–7. Ionospheric systems of electric currents which could pro-
duce the lunar daily magnetic variation L. The current-lines are
drawn at intervals of 1,000 amperes. (The maps are drawn similarly
to those in Figs. 7.15 and 7.16 for S; in the present case, the meridians
are numbered according to lunar hours)
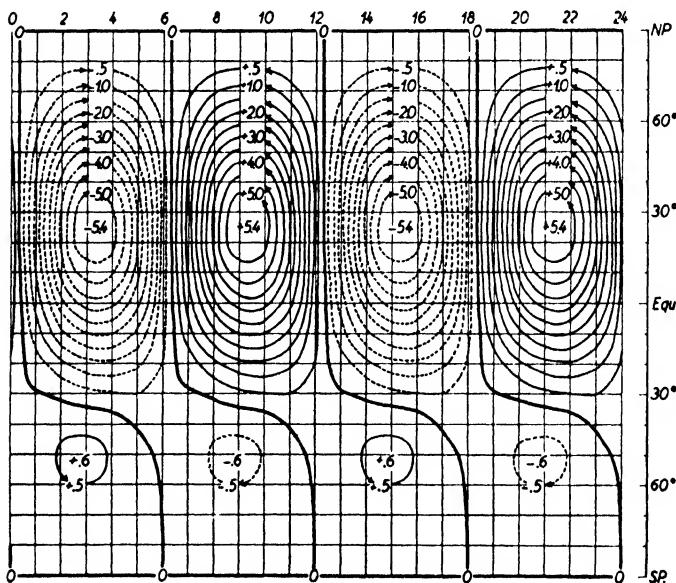
FIG. 5. Mean for a lunation, June solstice
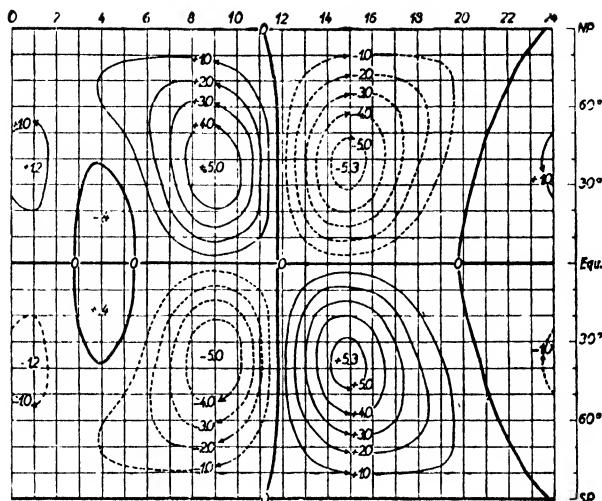
FIG. 6. New moon, equinoxes
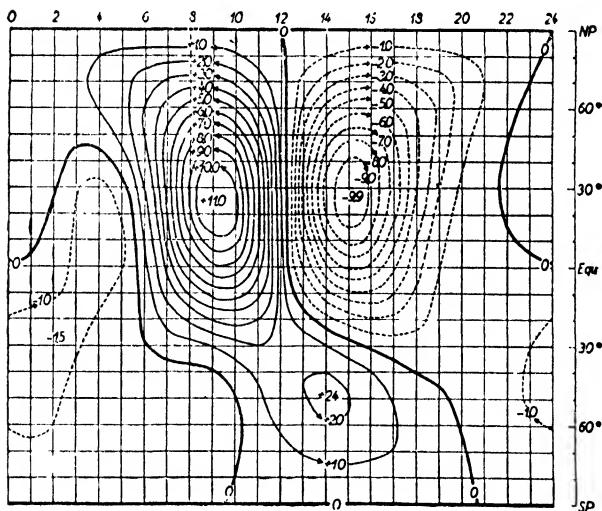
FIG. 7. New moon, June solstice

TABLE 8. *Provisional Harmonic Amplitudes for the 24-hourly and 12-hourly Waves in S and L for H at Huancayo, Peru, 1922–34*

Quiet and ordinary days (C < 1·2)

Season	No. of days	s_1	l_1	s_1/l_1	s_2	l_2	s_2/l_2
		γ	γ		γ	γ	
S. summer (Nov.–Feb.)	1,260	40·8	7·1	5·7	23·6	8·0	3·0
Equinoxes	1,261	51·8	4·3	12·0	28·5	5·9	4·8
S. winter (May–Aug.)	1,371	39·6	2·6	15·2	20·2	2·4	8·4

This table shows that the amplitude of L in H at Huancayo is comparable with that of S at some other stations. It is so great as to have
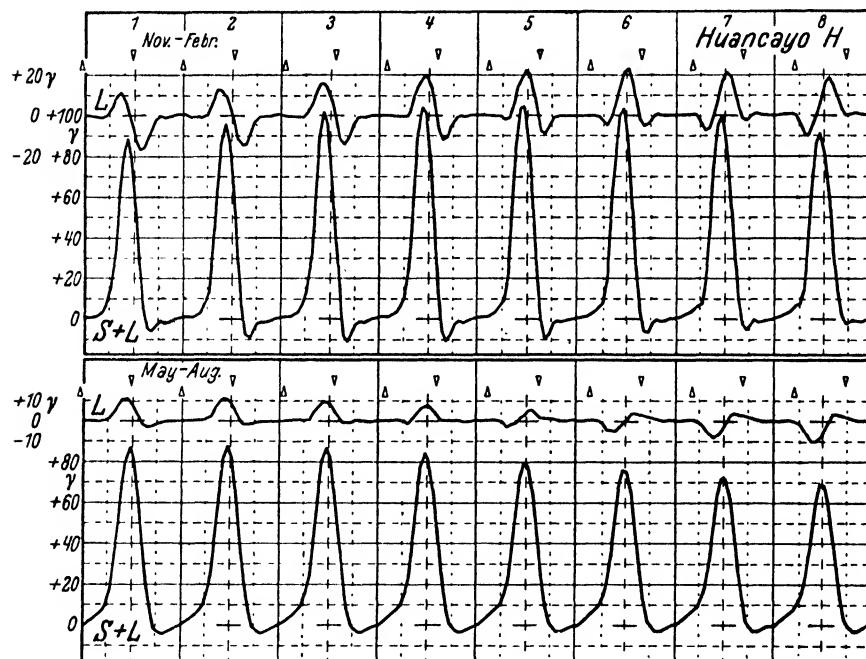
FIG. 8. The lunar daily (L) and combined solar and lunar daily (S+L) variations of the months November–February, or May–August. The

a clearly visible influence on the recorded curves for H [19a]. This is shown by Fig. 8, which gives curves for L, and also curves showing the combined S and L variations, for a succession of sixteen days (rather more than half a month); the fortnightly variation in S+L, due to the phase-changes of L relative to S, is evident. This curve is not an actual record but is constructed from a set of average daily S and L inequalities for groups of selected days (C < 1·2, daily sunspot number $s = 0$, for

the December solstice in the upper curves, and for the June solstice in the lower curves) at twelve different epochs in the lunar month; thus it is freed from the non-cyclic and fortuitous changes which affect the actual records. The upper and lower transits of the moon are marked on the figure. Fig. 10 shows actual records for two days, just before and somewhat after the new moon of 1933 December 16. The characteristic minimum in the evening on days after new moon (Fig. 9) is well shown on the actual record (Fig. 10).

The twelve daily inequalities for L, for twelve lunar phases, are illustrated in Fig. 11 (right), the parts corresponding to the hours of sunlight being thickened. The contrast between the L variation during periods of
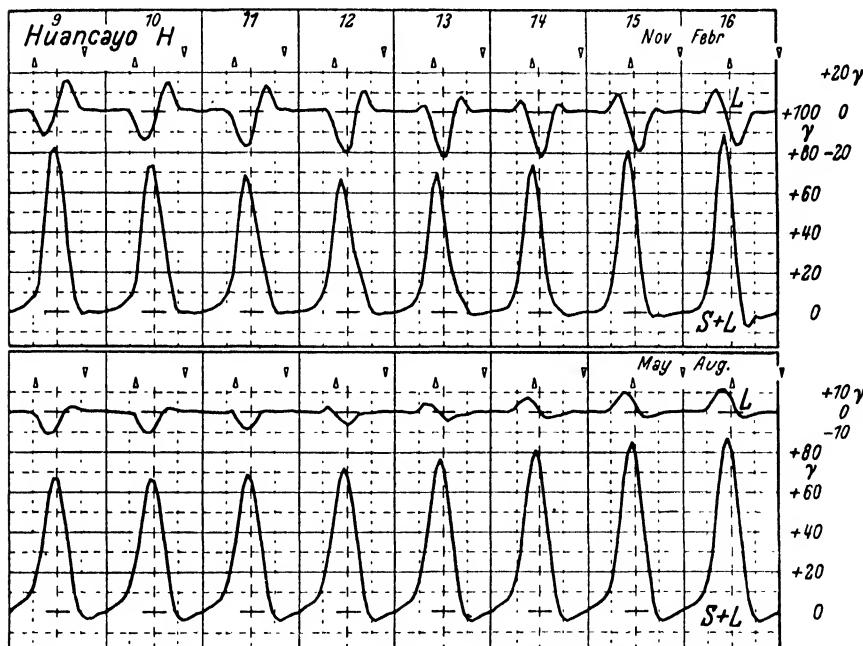
Huancayo horizontal intensity, on quiet days with spotless sun, during 16 days, in the upper and lower transits of the moon are marked

sunlight and darkness is very striking. The left part of Fig. 11, which shows S for the same days, on the same scale as for L, indicates that in S also the day-to-night contrast is unusually great; comparison of the left and right diagrams shows the relative magnitude of S and L in *H* at Huancayo.

The seasonal change of L at Huancayo is, as at other stations, much greater than for S. The December solstice being the period at which

L at Huancayo is greatest, the material for this period has been further examined. The days used in the first row of Table 8 were divided into two groups, according to the value of the daily sunspot number s; in one group $s < 30$ (mean $s = 8$), in the other $s \geqslant 30$ (mean $s = 72$).
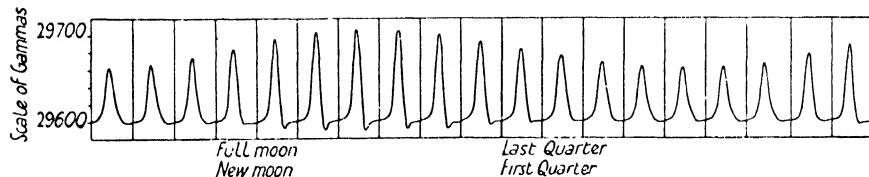
FIG. 9. The calculated combined curve for the solar plus lunar daily variation $S+L$ of horizontal intensity, during half a month of December, at Huancayo, Peru. Sunspot number zero, magnetic character-figure under 1·2
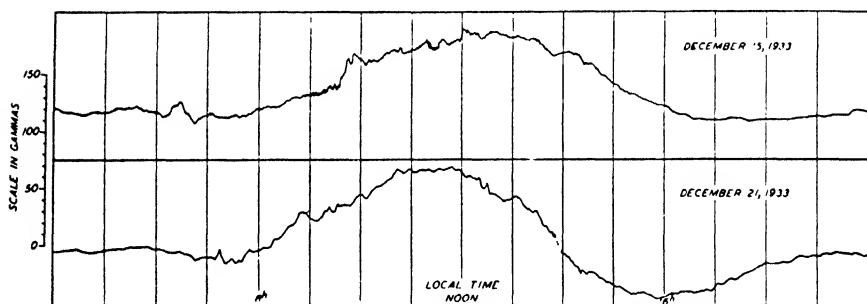
FIG. 10. Photographic records of horizontal force on December 15 and 21, 1933, at Huancayo, Peru, exhibiting the combined solar and lunar daily variation before and after the new moon which occurred on December 16

The results are given in the following table [19], together with those for the disturbed days ($C \geqslant 1·2$) of the December solstice, which were omitted in deriving Table 8, row 1.

TABLE 9. *Provisional Harmonic Amplitudes for the 24-hourly and 12-hourly Waves in* S *and* L *for* H *at Huancayo, December Solstice, 1922–34*

	No. of days	s_1	l_1	s_1/l_1	s_2	l_2	s_2/l_2
		γ	γ		γ	γ	
Few sunspots, $\bar{s} = 8$	715	33·6	6·8	4·9	21·2	7·6	2·8
Many ,, $\bar{s} = 72$	545	50·5	7·4	6·8	27·4	8·5	3·2
Disturbed days, $C \geqslant 1·2$	201	40·5	8·3	4·9	23·1	8·3	2·8
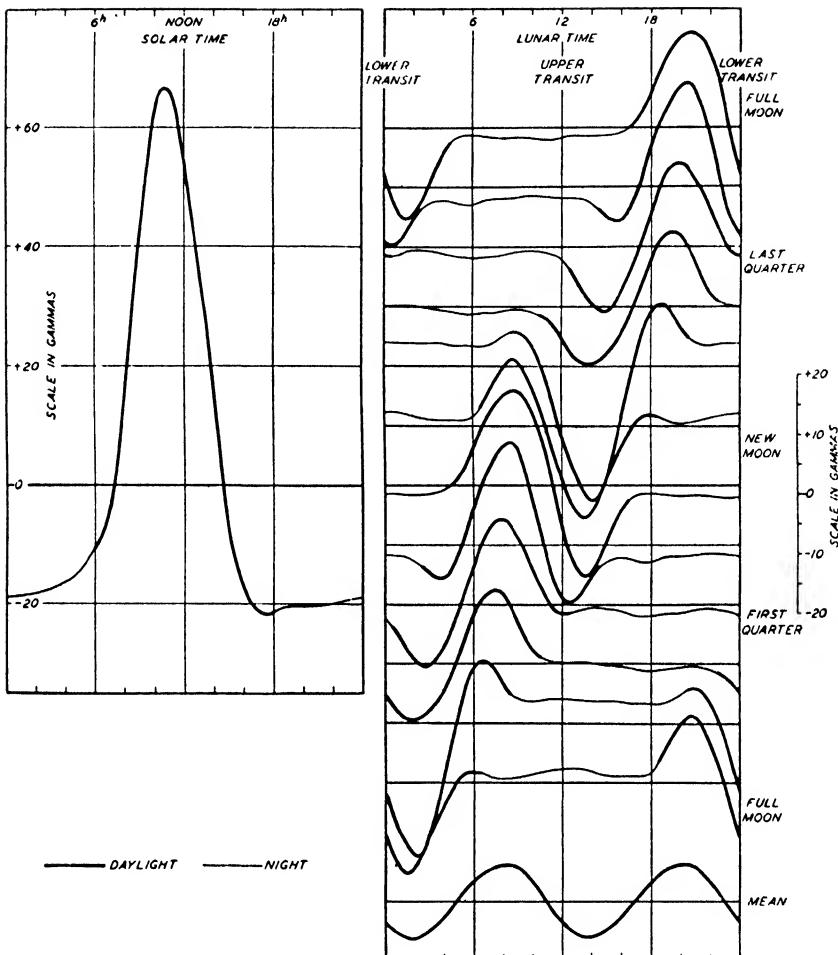
FIG. 11. Huancayo, Peru, horizontal intensity, December solstice, sunspot number zero, quiet and ordinary days. Left: The solar daily variation S. Right: The lunar daily variation L, at twelve phases of the moon (thick lines for hours of daylight) and for the mean of all phases

8.17. van der Stok's method. To find the change of L with the age of the moon, the method of Broun (§ 3), or an extension of it, has generally been used. A different method, introduced by van der Stok [18] and used by Figee [10] in his valuable memoir on L at Batavia, will be briefly described here.

Against each hourly value in the data, the mean lunar hour (to the nearest integer) was written. The values at a given solar hour (say 2^h)

were then copied out, those for the same lunar hour being put in the same column; by averaging in each column, a lunar daily inequality is obtained, derived wholly from values for the selected solar hour. Twenty-four such inequalities are obtained, one for each solar hour: those for the solar hours of daylight have a much larger amplitude than those for night hours. For example, for Batavia D, in the half-year October–March, the average deviation from the mean, in the lunar daily inequalities for the 8 solar hours centred at noon, was 0·285′, whereas it was only 0·045′ for the 8 solar hours centred at midnight.

The results of this method of rearranging the hourly values must, of course, be essentially the same as those obtained by Broun's method; but the latter seems preferable, because when it is used the hourly values contributing to any lunar daily inequality are much more consecutive, and are derived from fewer days, than in van der Stok's method; hence they are less influenced by changes from day to day, which must particularly affect van der Stok's inequalities when these are derived not from all days over a period of months or years, but from selected days over such a period—days of small sunspot number, or of high magnetic activity, or maximum lunar distance—which may follow one another at irregular intervals of several days.

8.18. The influence of the exclusion of disturbed hours. In applying van der Stok's method to D at Batavia, the hourly values used by Figee were hourly deviations from the daily mean, from which the solar daily variation S had been removed. Thus they represented L+D. He included all hours, disturbed or quiet, in forming his lunar daily inequalities, from which most of D was averaged out.

In a separate investigation [10], however, he examined whether or not the lunar hourly departures from the mean would have been modified if he had excluded the more disturbed hours, for which the hourly deviation (plus or minus) was numerically greatest. At the time of his work, and earlier, it was usual to exclude such hours in the determination not only of L but also of S; Sabine, for example, advocated this practice [14].

The nature of his results may be illustrated by a particular example, referring to the averaged deviation at the lunar hour 0, in the lunar daily inequalities for all days, at all ages of the moon, in the four months November–February (southern summer) for the years 1884–99. Including all the hourly deviations, the average was 0·22′; when those numerically in excess of 1·2′ were excluded, the average was 0·17′.

This reduction by exclusion of the disturbed hours (unless it is merely accidental) is sufficiently large to justify Figee's conclusion that the disturbed hours should not be excluded.

Figee also computed the average of the deviations for the excluded hours, and found it to be 1·05′—nearly five times as great as the
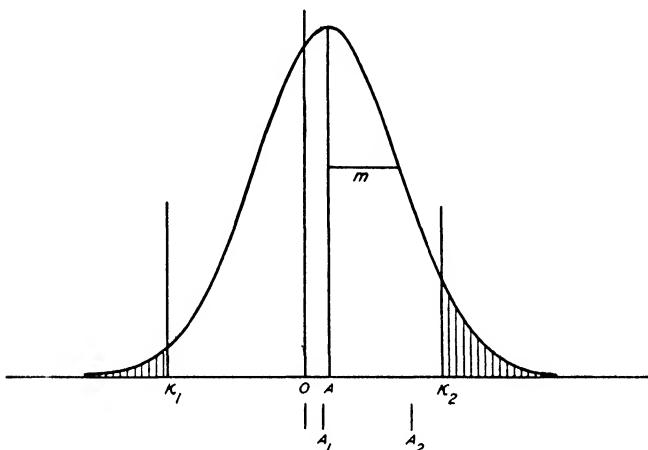
FIG. 12. The statistical effect of excluding large numerical values, symmetrical about an origin O, from a distribution centred at a point A differing from O

average; it is because of this large average that the relatively few excluded hours (110 out of 1,774) could so appreciably affect the mean.

Though Figee himself did not conclude from this result that L was five times as great during these disturbed hours as on the average, his work seems at times to have been interpreted in this sense. As the fallacy is occasionally met with also in other geomagnetic studies, it seems worth while to expose it definitely here, by a simple idealized example.

Consider a set of numbers x whose frequency distribution about their mean value x_0 accords with the normal error law with standard deviation m (16 (67)), as in Fig. 12, where $OA = x_0$. Let x_1 ($= OA_1$ in the figure) be the mean of the set after excluding the numbers numerically greater than b ($= OK_1 = OK_2$ in the figure). Since the frequency curve is symmetrical about the ordinate through A, the excluded parts are of unequal size, that beyond K_2 exceeding that beyond K_1. The mean x for these excluded parts is represented in the figure by OA_2 or x_2, and much exceeds OA_1. If fractions n_1, n_2 of the whole area under the

frequency curve lie respectively inside and beyond the interval $K_1 K_2$, so that $n_1+n_2 = 1$, then

$$n_1 x_1+n_2 x_2 = (n_1+n_2)x_0,$$

or $$(x_0-x_1)/n_2 = (x_2-x_0)/n_1,$$

or $$AA_1/n_2 = A_2 A/n_1.$$

Thus, if n_1/n_2 is large, the mean x_2 will much exceed x_1. In the figure this is illustrated for a case analogous to that of Figee's data: $m = 0.65'$, $b = 1.92m = 1.18'$, $x_0 = 0.32m$, $x_1 = 0.24m$, $x_2 = 1.48m$.

This makes it clear that the result obtained by Figee has no immediate bearing on the physical question whether L is larger during disturbed than during quiet times.

8.19. Other lunar geomagnetic effects. Attempts have been made [16.2, 3] to find other lunar influences on the earth's field, such as a difference between the daily mean at new moon and at first quarter. Apart from a slight effect in Huancayo H, no significant change appears to have been established as yet, and the search is made difficult by the interference of the monthly interval with the 27-day recurrence phenomenon due to solar rotation (Chs. XII and XVI).

8.20. Newer methods for the determination of L. The Huancayo results given in § 16 were obtained by Bartels [19] and Schneider [20] by harmonically analysing each solar daily sequence of bi-hourly values of H, and plotting the results on a harmonic dial (·16.18); the mean of the points P_2 whose abscissae and ordinates are (for example) a_2 and b_2 gives a_2 and b_2 for the mean semi-diurnal component of S. The deviations of the individual points P_2 from the mean point P_2' are due to L and to accidental variations. By plotting these deviations $P_2' P_2$ on a new diagram, on which each $P_2' P_2$ is turned through a suitable angle so as to bring these deviations from days at different lunar ages into the same phase relative to the moon, we obtain a new set of points P_2'' (namely, the ends of these orientated lines $P_2' P_2$ all drawn from a common origin O representing P_2'); the mean of the points P_2'' gives a_2 and b_2 for L.

Miller and Chapman have devised suitable analytical methods for determining L from large masses of data. Their method is applicable to hourly, bi-hourly, or tri-hourly data, but in practice they follow Bartels in the use of bi-hourly data. An important feature of their method is the allowance made for the non-cyclic variation of the magnetic element in the course of each day. Days at the same lunar ages (in twelve groups) are grouped together, and the bi-hourly sums are

analysed harmonically, giving twelve sets of values a_n, b_n ($n = 1$ to $n = 4$). The harmonic components of L are derived from these by further harmonic analyses of the sequences a_n and b_n for the twelve groups of days, and of the sequence giving the number of days in each group. An account of this method is shortly to be published. A considerable mass of results obtained by this method (or earlier methods) now awaits publication; it includes determinations of L for Greenwich H, for Cheltenham (U.S.A.) (all elements), for Helwan (all elements), and for Sitka D.

THE MORPHOLOGY OF MAGNETIC DISTURBANCE

9.1. Separation of the storm-time and the disturbance-daily variations.

The study of magnetic disturbance is perhaps best approached by consideration of the outstanding disturbances called magnetic storms. The most intense storms appear to start suddenly at the same time, to within about a minute, all over the earth: their decay and disappearance are less definite. Often there are remarkable similarities as well as great differences between the effects of the storm on the variations of different elements at different observatories. Fig. 1 illustrates this [1]; it refers to part of the storm of 1885 June 24, 25.

It has long been known (Broun [1 a]) that for some time after a period of great disturbance H is reduced below its mean value. This information is rendered much more complete by a procedure first applied by Moos [G 46] to Bombay records of storms. In the period 1872–1904 he found 113 storms having sudden commencements (S.C.s): such storms may be called S.C. storms (Moos called them X storms); their initial times were distributed fairly evenly throughout the 24 hours. He determined the average departure of H during these storms from its mean value in the corresponding months, at each hour up to 24 after the commencement of the storm; it is convenient to refer to time so measured as *storm-time*. He found that the results gave a smooth curve, showing a rise of 26γ in H, during the first hour, above the initial value, then a rapid fall until at 10^{h} (storm-time) H attained a minimum, about 86γ below its initial value; this was followed by a slow recovery towards the mean; in the remaining 14 hours of the first day the recovery amounted only to 24γ. Subdivision of the data into three groups for three 11-year periods gave similar curves, though the irregularities were not averaged out so completely. Similar curves were obtained on examining the 134 greatest storms of the period 1872–1904 (half of them being included in the 113 S.C. storms already referred to). For these 134 storms the storm-time changes in D and Z were likewise examined, and were found to be small compared with those in H.

Further, after abstracting these average changes, proceeding according to storm-time, from the variations of D, H, and Z during the first day of the storm, the hourly differences were rearranged according to local time, and daily variations were derived: from these the ordinary

monthly mean daily variations were subtracted, so that the remainder showed the effect of disturbance in modifying the average solar daily variation. The additional variation may be called a disturbance-daily variation: it would be S_D, as defined in 6.8, if S_q instead of S had been
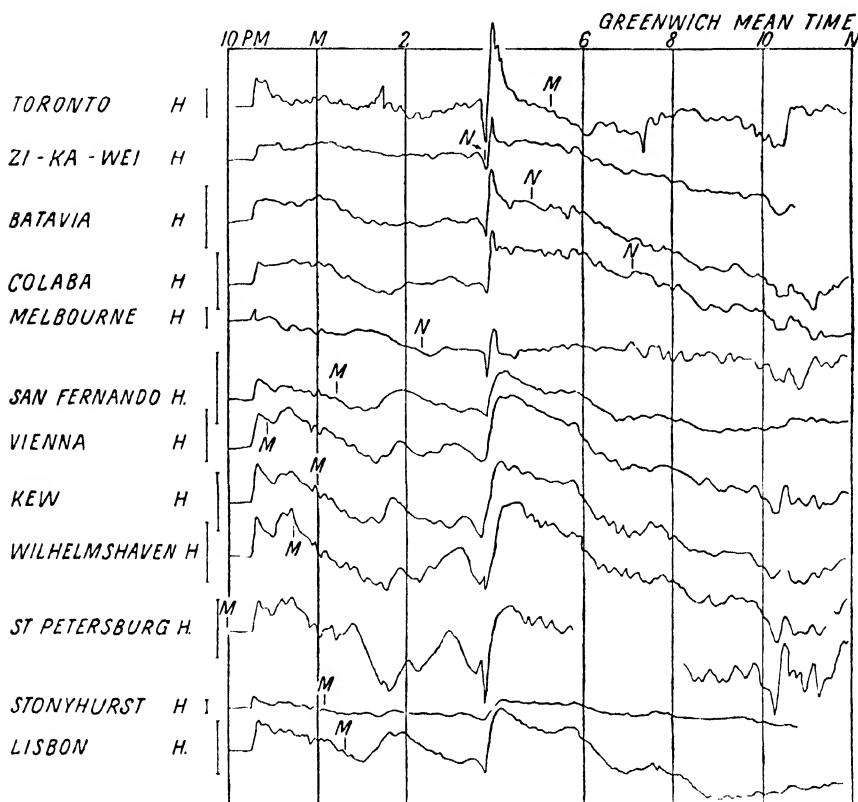
FIG. 1. Records of the variation of the horizontal force during the magnetic storm of 1885 June 24, 25. (After W. G. Adams.) The scales are indicated by the lengths of the vertical lines marked H, which correspond to 100γ or 0·001 c.g.s. Two changes, at 10.32 p.m. and 3.48 a.m. were simultaneous over a great part of the earth. Local noon and midnight are indicated by N and M

removed; actually it is only a fraction of S_D, but it is convenient to ignore the difference, and to call the additional daily variation, so formed, by the name S_D. This was found for D, H, and Z from the 134 greatest storms, and for H also from the 113 S.C. storms: the two sets of results for H agreed well together.

Similar methods were applied to the second 24 hours of 29 selected intense storms; they showed that the rise in H, proceeding according to storm-time, which marked the latter part of the first day of the

storms, was continued throughout the second day at a diminishing rate: the S_D variations were also determined, but the results obtained for these [G 46, § 446, p. 469] are not confirmed by later investigation. Moos also examined the records of (only) seven storms at Kew in a similar manner (for the first 24 hours).
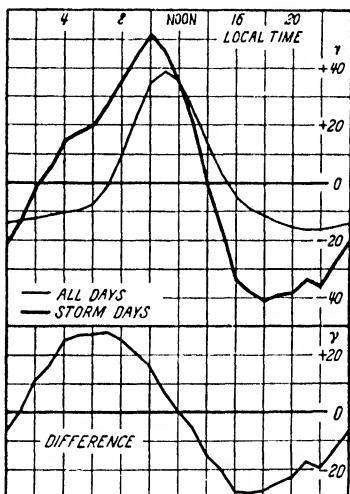
FIG. 2. The solar daily variations of horizontal intensity at Bombay; averages from all days and from storm days
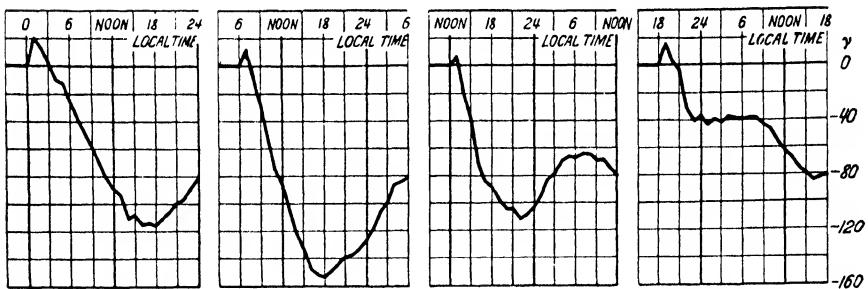
FIG. 3. The changes of horizontal intensity at Bombay during magnetic storms for different local times of storm commencement

Fig. 2 is an example of one of Moos's curves, showing the S curve for H at Bombay for storm days, all days, and storm days minus all days; the latter approximately represents S_D. This should be compared with the curves 5 a (all days) and 5 b (storm days minus all days) in

Fig. 6 below. Examples of storm-time curves for Bombay and other stations are given in Figs. 5 and 4 below.

Fig. 3 gives the combination of the average storm-time variation D_{st} and the average S_D variation, derived from a group of Bombay storms, for different local times of storm commencement. It is an idealization of actual storm curves in that most of the irregular variations D_i have been removed in the averaging referred to. It shows how the local time of commencement affects the course of development of a typical storm.

9.2. The storm-time variation. Chapman [2] applied the method used by Moos, to data from many observatories, over a considerable range of latitude. Fig. 4 shows the storm-time changes in H, D, and Z for groups of stations as follows:

(1) Batavia, Porto Rico, Honolulu (22°).

(2) Zikawei, San Fernando, Cheltenham, Baldwin (40°).

(3) Pola, Potsdam, Greenwich (51°).

The numbers in brackets give the mean magnetic latitude of the group. Fig. 4 was derived from forty selected storms of fairly uniform moderate intensity, the initial times being well distributed over the day. The scales of time and force are indicated at the bottom and the sides of the diagram. Since Batavia is a southern station, where the changes in Z and D are opposite to those in northern latitudes, the variations in these elements were reversed in combining them with those for other stations in the first group.

Fig. 4 shows that the initial rise and subsequent larger decrease in H, followed by slow recovery, occurs in all latitudes over the range considered, but that the maximum diminution decreases with increasing latitude. In the higher latitudes the irregular variations increase relatively to the average storm-time variations, and render the latter more difficult to determine.

The storm-time variation in the magnitude of Z is opposite to that in H, and much smaller in magnitude: also, unlike that in H, it is reversed in the southern hemisphere. There is practically no storm-time variation in D.

The upper curves in each section of Fig. 5 illustrate the storm-time variation in H at Bombay and Pavlovsk for three groups of storms selected according to their intensity. For Pavlovsk the results are given for the first 48 hours of the storm. The general character of the changes is the same in the three groups of storms, but the epoch of minimum
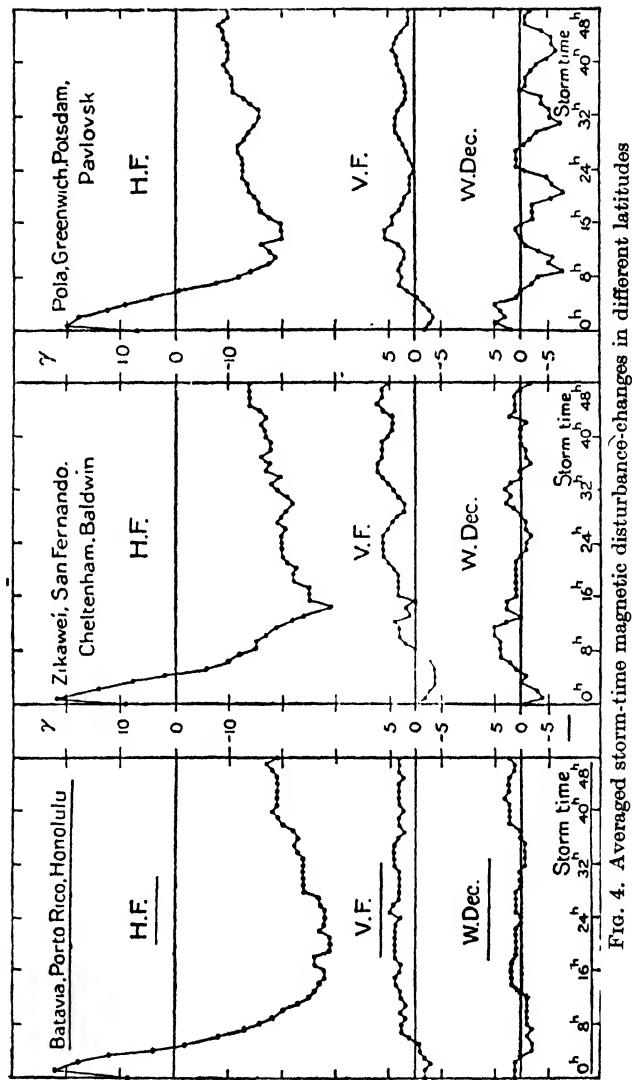
Fig. 4. Averaged storm-time magnetic disturbance-changes in different latitudes

H is attained notably sooner in the more intense than in the moderate storms, and, in all groups, rather earlier for Bombay than for Pavlovsk.
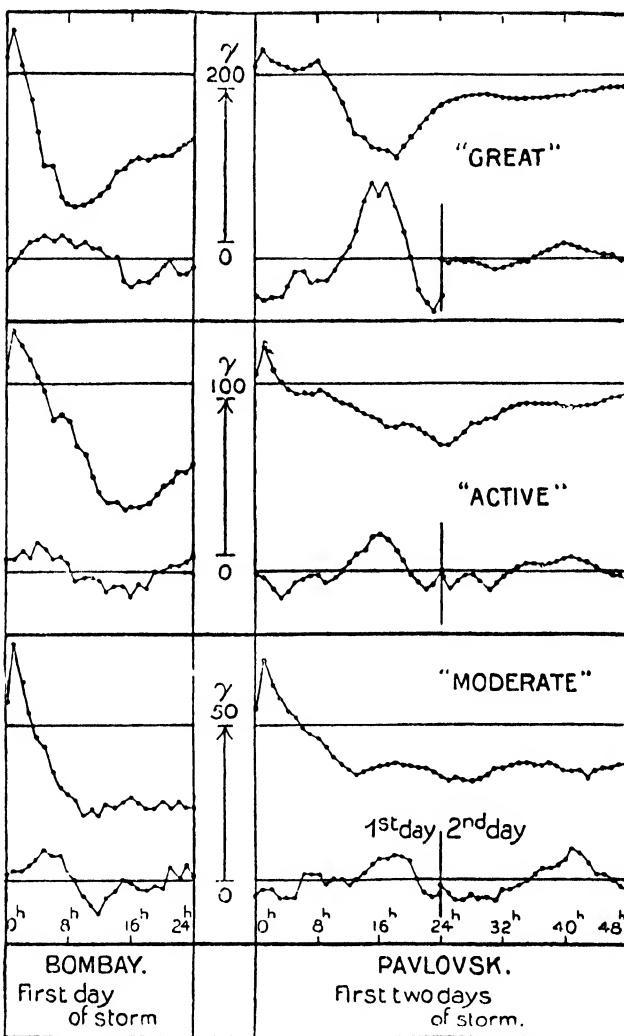
Fig. 5. Averaged storm-time and disturbance-daily variations for groups of magnetic storms of different intensities

9.3. The disturbance-daily variation S_D.

Figs. 6, 7, 8 (columns b, c) show the disturbance-daily variation in H, Z, and D, proceeding according to local time, for the first and second days from the commencement of the forty moderate magnetic storms illustrated in Fig. 4. Column a shows the average monthly mean daily variations in the three
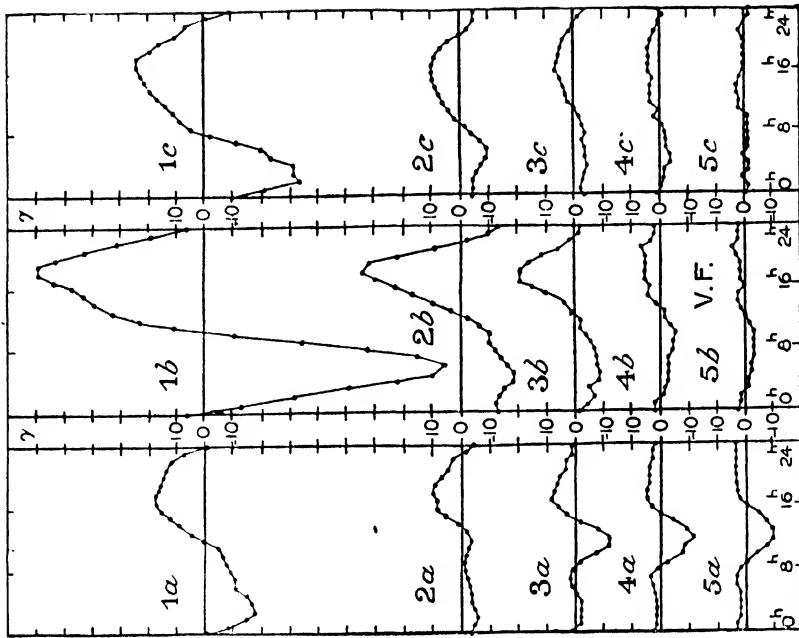
Fig. 7. Quiet (a) and disturbance (b, c) daily variations in vertical force Z; b, c refer to the first and second days of magnetic storms
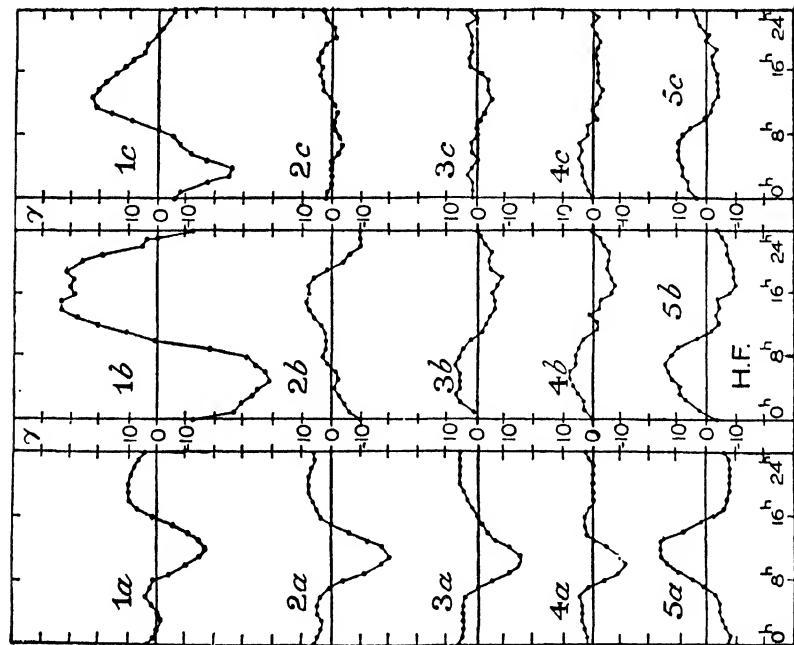
Fig. 6. Quiet (a) and disturbance (b, c) daily variations in horizontal force (H.F.); b, c refer to the first and second days of magnetic storms

elements for all days of the months of the forty storms considered; the curves a are in most cases mainly due to S_q, but also include a fraction due to the disturbance-daily variation S_D.

The curves (b, c) in Figs. 6–8 represent the daily variations on the storm days, less the average daily variation S represented by the curves (a); here S itself consists partly of S_q and partly of S_D, so that while S_q has been removed from the curves (b, c), so also has part of the S_D variation. Hence these curves represent most, but not quite all, of the S_D variations on these days of storm.

In Figs. 6–8 five sets of curves 1–5 are given for Sitka (mag. latitude 61°), Pavlovsk (58°), Pola, Potsdam, and Greenwich combined (51°), and for the groups shown in the centre (40°) and at the left (22°) of Fig. 4. (The lower curves in each section of Fig. 5 also refer to S_D.)

It is clear that the S_q variation (represented fairly well, though with a certain admixture of S_D, by the curves (a)), is quite different in type from the S_D variation. For example, in Fig. 6 there is a reversal of type in section (a) in a low latitude (about 30°, between the latitudes to which the curves (4) and (5) refer), while in sections (b) and (c), though there is also a reversal of type, it occurs in a much higher latitude (about 55°, between the latitudes to which the curves (2) and (3) refer; moreover, in section (a) the extreme departures from the mean occur in the middle of the day, at about the time when the curves (b) and (c) pass through the zero-line. The latter contrast is shown also by Fig. 7, though in other respects this differs greatly from Fig. 6. In Fig. 8 the curves (a) pass through zero about midday, while the curves (b) and (c) pass through zero at about 6^h and 18^h. The principal difference between the columns (b) and (c) in all the three figures 6–8 is in the amplitude, indicating a decline in S_D after the first day of the storm.

Examination of the local-time variation derived from the first hour of storm-time (when H is above its initial value) for a number of Greenwich storms commencing at various hours of the day, shows that S_D in H is then similar to that illustrated in Fig. 6, but opposite in phase.

Perhaps the most remarkable feature of Figs. 6–8 is the great increase of S_D in Z on passing from the equator to Pavlovsk and Sitka. The curves Fig. 7 (a) for Sitka and Pavlovsk may therefore be expected to include a relatively greater amount of the S_D variation than the other curves (a); and their appearance supports this expectation.

The results of Figs. 4–8 indicate that up to latitude 60° (roughly)

the average disturbance field during magnetic storms has a well-marked character. Figs. 6–8 show this as regards distribution in latitude and in local time (or in longitude relative to the sun); Figs. 4, 5 show it as regards time measured from the commencement of the storm; it may
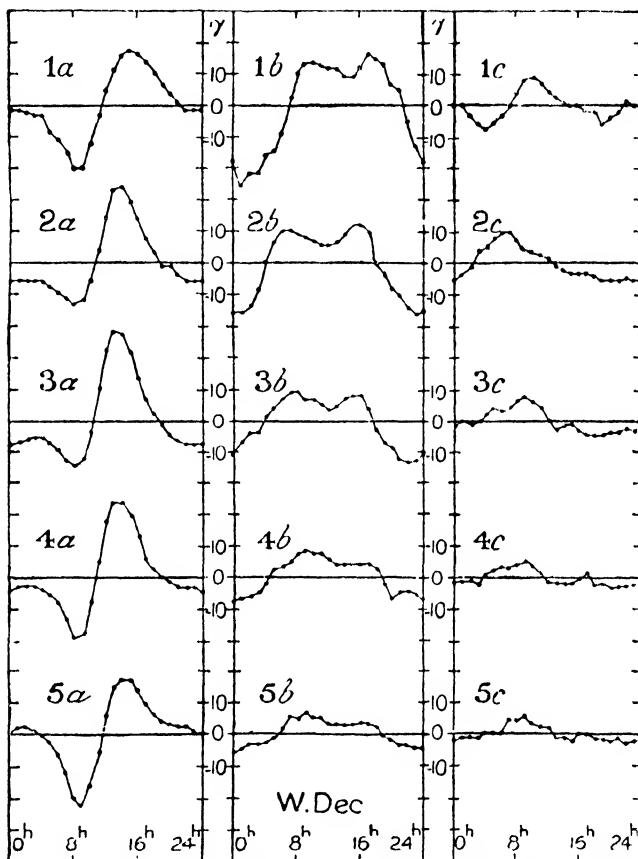
Fig. 8. Quiet (a) and disturbance (b, c) daily variations in West declination; b, c refer to the first and second days of magnetic storms

be noted that the storm-time variations correspond to a part of the disturbance field which is independent of longitude. Fig. 5 suggests that the general character of the field and its changes is the same over a fourfold range of intensity, though the rate of progression of the field, through its successive phases, depends to some extent on the intensity of the storm. It should be added that, in addition to these average variations, there are, of course, less regular features peculiar to each individual storm.
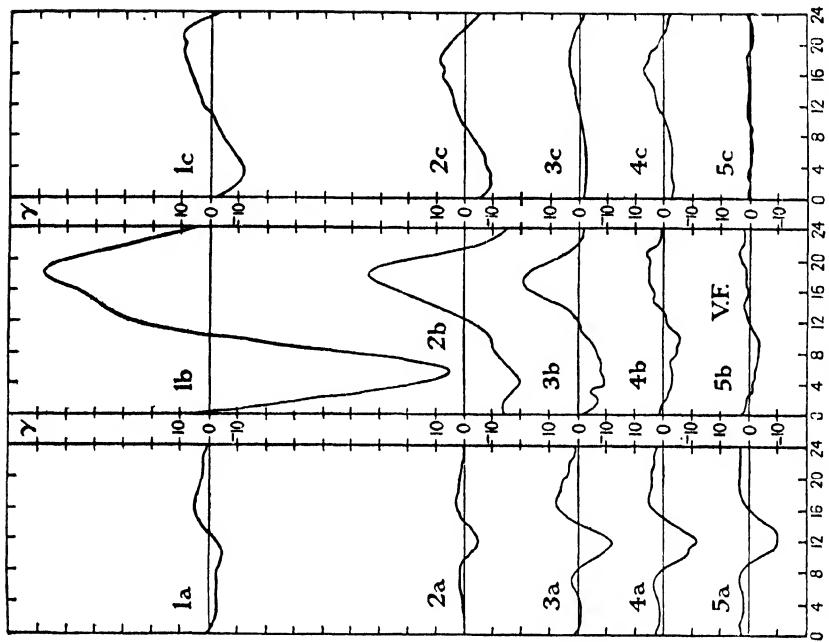
Fig. 10. Quiet (a) and disturbance (b, c) daily variations in vertical force Z; b refers to storm days and c to average days
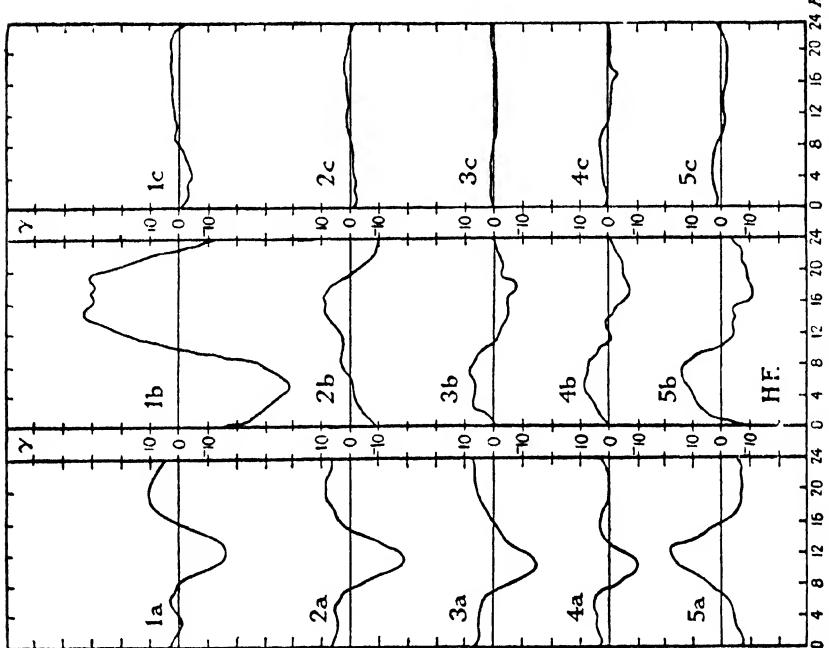
Fig. 9. Quiet (a) and disturbance (b, c) daily variations in horizontal force (H.F.); b refers to storm days and c to average days

9.4. The disturbance-daily variation for slight and intense disturbance.

It is important to ascertain whether the *slight* magnetic disturbance which distinguishes the average day from specially quiet days differs from *intense* disturbance (or storms) not only in degree but also in type. The simplest basis of comparison is provided by the S_D part of the D field, which is illustrated in Figs. 9–11. These are the same as Figs. 6–8 in columns (*a*) and (*b*), but in column (*c*) they show S_D corresponding to the slight amount of the D field present on average days; the curves therefore represent the difference between the daily variations of *H*, *Z*, and west *D* as derived from all days (in one or more years), and as derived from quiet days. The (*c*) curves 1–5 refer to Sitka, Pavlovsk, Greenwich, Cheltenham, and Honolulu, respectively, and so correspond to much the same latitudes as the curves 1–5 (*a, b*). Comparison of the curves (*c*) of Figs. 9–11 with the curves (*b*) makes it evident that S_D preserves essentially the same character while the intensity of disturbance ranges from storms down to the low limit corresponding to that present on ordinary days. The fifteen independent curves (*b*) of Figs. 9–11, in spite of minor irregularities (many of which would disappear if averages of a larger amount of material were taken), are very similar to the corresponding curves (*c*). The similarity of the curves (*c*) to the curves (*b*) of Figs. 6–8 bears on the same point.

Figs. 6–11 summarize a large amount of evidence, which can be amplified, for example, by considering the seasons separately. This has been done recently by Cynk [40], who confirms the general conclusions stated above, and also finds the seasonal variations of the S_D field.

9.5. The storm-time variation for slight and intense disturbance.

The evidence of § 4, indicating the similarity between strong and weak disturbance, relates only to one part of the additional magnetic field D present during disturbance. The other systematic part depends on storm-time, and represents the effect of the disturbance, at successive hours after the commencement, averaged round each parallel of latitude. Unfortunately the similarity of type as between intense and minor disturbance cannot be tested, in regard to the storm-time variation, in the same detailed way as is possible for the local-time variations. The difficulty lies in the determination of storm-time, reckoned from the beginning of the disturbance, in the case of weak disturbance. It is often easy to judge the time of commencement in the case of magnetic storms, because they are well-marked individual phenomena. But it is usually hard to distinguish whether a prolonged period of minor disturbance is made up of one or of many individual

disturbances. It is possible, however, to apply a partial test of the similarity of type for strong or weak disturbance, also in the case of the storm-time changes.

The local-time changes considered in § 4 are measured from the daily
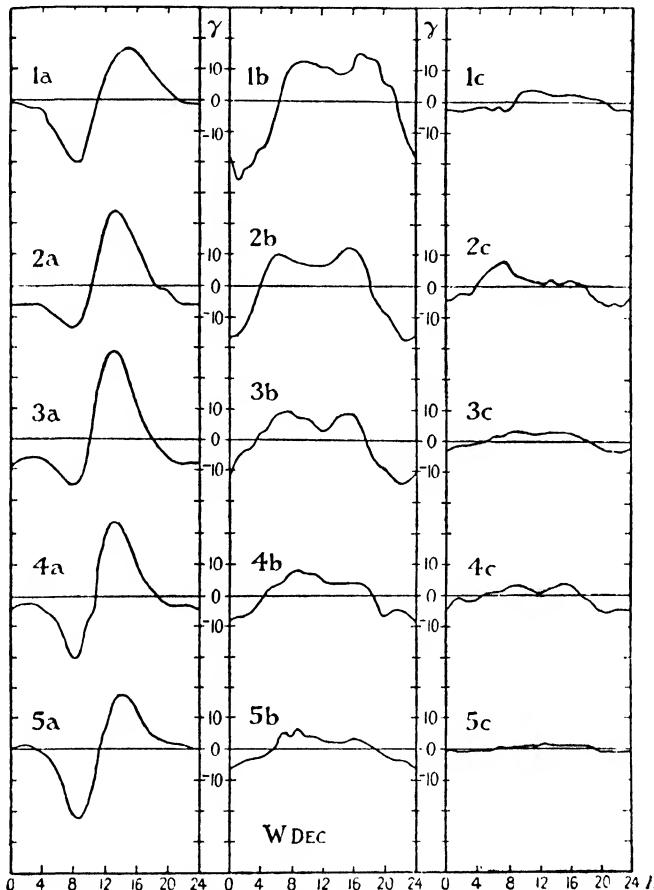
FIG. 11. Quiet (*a*) and disturbance (*b*, *c*) daily variation in west declination; *b* refers to storm days and *c* to average days

mean value of the magnetic element as origin, and are independent of the absolute value of this mean. The storm-time changes, on the other hand, directly affect the daily mean value, and the way in which they will do so on ordinary days, if they are then of the same type as in magnetic storms, can be inferred on the reasonable assumption that, when a large number of more or less disturbed days are considered, each hour of the day will experience approximately the same *mean* storm-time effect D_m. For example, during the first few hours of a magnetic

storm the horizontal force is generally increased (taking the average round a parallel of latitude), but subsequently it experiences a much larger and more prolonged decrease, which dies away slowly: thus the average storm-time effect on the horizontal force should be a decrease in the daily mean during disturbance. The magnitude of the decrease should have a maximum at the equator (cf. Fig. 4), and should diminish with increasing latitude to about half the equatorial value, at about 50° latitude.

The vertical force storm-time changes are much smaller, and opposite in sign, but otherwise follow a similar course. The averaged effect on the daily mean value of the downward vertical force should be a very slight increase, in moderate northern latitudes: the effect should vanish at the equator and be reversed (like the whole vertical force) in the southern hemisphere.

The west declination shows no appreciable average storm-time changes in lower and middle latitudes, and therefore the daily mean value of the declination (more exactly, of the force component perpendicular to the plane through the earth's magnetic axis; see Fig. 19) should not be affected by disturbance. In individual storms, of course, the disturbance in declination may be very great (see § 22).

These expectations, based on the conclusion of § 4 that the average characteristics of slight and intense disturbance are similar in type, are well borne out by observation, as the following table shows. It gives the differences obtained by subtracting the daily mean values of the three magnetic elements for *quiet* days from the corresponding daily means for *all* days. The stations are those already considered in § 4 (their magnetic latitudes are indicated in brackets): the results are expressed in terms of the unit 1γ.

	H.F.	V.F.	W. Dec.
(1) Sitka (61°)	−3	−3	0
(2) Pavlovsk (58°)	−5	0	0
(3) Greenwich (55°)	−4	0	0
(4) Cheltenham (49°)	−7	1	0
(5) Honolulu (21°)	−10	1	1

Though these results do not all refer to the same series of years, a different choice would not alter the general indication of this table. The changes observed are small except in the case of the horizontal force, where they are systematic, and in good agreement with expectation. The results are given to the nearest unit, and are probably accurate to within about 1γ.

In addition to this evidence favouring the view that also as regards the storm-time changes slight and intense disturbances are similar in type, there is the direct evidence afforded in Fig. 5, showing that the type of the storm-time changes during intense disturbance is but little altered over a range of intensity varying in the ratio 1 to 4. The progression through the storm-time changes seems to be quicker the more intense the disturbance, but the similarity of type is the outstanding feature.

Thus it appears that, in addition to the main magnetic field of the earth and the varying field which is manifested by the daily magnetic variations on quiet days, there exists, from time to time, a further 'disturbing' magnetic field D, which has definite average characteristics in regard to its space distribution and temporal changes; and that, over a wide range of variation of intensity of this field, these average characteristics do not vary greatly in *type*.

This conclusion relates to world-wide magnetic disturbance and not to isolated local or brief disturbances (such as 'bays'), except in so far as these are associated with, and indicative of, the general disturbance which distinguishes ordinary from quiet days.

DISTURBANCE IN HIGH LATITUDES (§§ 6–9)

9.6. Storm-time changes in high latitudes. The evidence so far described relates to latitudes up to about 60°: the data for polar regions will next be considered. The observational material is more limited than for lower latitudes, particularly as regards the period of continuous photographic registration, which at few polar stations has exceeded two years. Consequently no diagrams for any polar station, analogous to Figs. 4–8 for lower latitudes, have yet been prepared from forty or more storm days.

In particular, the course of the storm-time changes in the polar regions has not yet been determined. But it is possible to compare days of different degrees of slight disturbance (reckoned by polar standards), as was done in §§ 3, 4 for non-polar stations; that is, the S_D variations and the averaged storm-time variations can be found for polar latitudes. The results so obtained will, of course, have reference to the same relatively low degree of disturbance as in the case of the curves (c) in Figs. 9–11.

The averaged effect of the storm-time changes will first be considered, as shown by the difference between the daily mean values of the magnetic elements on all days and on quiet days. The data used are those

obtained at several Arctic stations during the international expeditions of 1882–3, and the results of later Antarctic expeditions, and also the hourly values for 1914 and 1915 recorded at the permanent Arctic observatory at Sodankyla, in Finland, a little to the south of the auroral zone. Where the published observations include results for quiet days separate from those for all days, these two sets of results have been used; where quiet-day results are not quoted, they have been computed for sets of quiet days chosen by the criterion of relatively small daily range. The results so obtained may not be strictly comparable *inter se*, but suffice for the present purpose.

The difference between the all-day and quiet-day means for each station is a vector with three components: in polar regions the horizontal component of this vector does not in general lie so nearly along the direction of the whole horizontal force as it does in lower latitudes, and, moreover, the geographical direction of the horizontal force varies considerably. It is, therefore, desirable to indicate graphically the direction of the horizontal disturbance vector, as in Fig. 12.

In this figure the situation in latitude and longitude is represented for ten observatories by a point in each case, and from each point a line (having an arrow-head at the farther end) is drawn indicating in magnitude and direction the difference between the mean horizontal component of magnetic force on all days and on quiet days. The zone of maximum auroral frequency, as drawn by Fritz, is also shown (dotted). It appears that, on the whole, these horizontal disturbance vectors diverge from a point or small region near the centre of the auroral zone (this point is also approximately the pole of the axis of magnetization of the earth); their distribution is *not* symmetrical about the geographical pole.

The change in the mean horizontal force thus consists approximately of a *reduction* in the component of the horizontal force along the meridians through the magnetic axis, and in this respect it resembles the corresponding horizontal-force change in lower latitudes (§ 5). But the *magnitude* of the reduction is larger in polar regions (where it ranges up to about 20γ) than in lower latitudes (where the maximum, which occurs at the equator, is about 10γ, though these figures may not be exactly comparable, owing to the different years and days from which they are drawn).

The average storm-time change of force in the horizontal plane is thus everywhere a decrease; its numerical value has a maximum at the equator, whence it decreases towards a minimum at about 60° latitude,

afterwards increasing rapidly towards the auroral zone. Its variation, and that of the vertical force within the zone, has recently been studied by Vestine and Chapman [20], using 1932–3 data.

The corresponding changes in the vertical force are more difficult to ascertain, since not all the polar stations of 1882–3 were equipped with reliable (or, in some cases, with any) vertical-force magnetographs. But
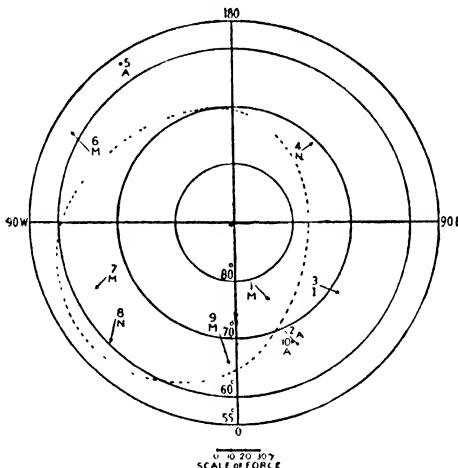
FIG. 12.   All-day minus quiet-day mean horizontal-force vectors in high latitudes. 1. Cap Thordsen. 2. Bossekop. 3. Nova Zembla. 4. Ssagastyr. 5. Sitka. 6. Fort Rae. 7. Kingua Fjord. 8. Godthaab. 9. Jan Mayen. 10. Sodankyla

M = Morning V.F. maximum   A = Afternoon V.F. maximum
I = V.F. curve irregular   N = V.F. record missing or unreliable

examination of the available data makes it appear that at Fort Rae, Nova Zembla, and Kingua Fjord the all-day mean is distinctly higher than the quiet-day mean, by about 18γ, 18γ, and 6γ respectively. At Bossekop, Sodankyla, and Sitka the change is negative and small, viz. -4, -4, -3γ. At Cap Thordsen and Jan Mayen the change appears to be small.

The large changes of vertical force, for which numerical values have just been quoted, are probably reliable at least as regards sign and order of magnitude; they much exceed the corresponding vertical-force changes in lower latitudes, which are small and positive from the equator (where the change is zero) up to about 55° northern latitude; there they change sign (to negative) and increase numerically towards the auroral zone. A further change of sign, to positive values, occurs at or within the auroral zone, where the large differences above noted are found. The recent study of the 1932–3 data by Vestine and

Chapman [20] shows that farther within the zone the differences decrease again towards the axis of magnetization.

In the Antarctic the results for Cape Evans (77° 38′ S., 166° 24′ E., geographical), reduced and discussed by Chree [3], show that the 'all-day minus quiet-day' mean difference in the (upward) vertical force is an increase of 5γ; this result is confirmed by the still larger difference of 12γ, with the same sign, shown by the 'disturbed-day minus all-day' means (from the five most disturbed days per month). Thus within the southern auroral zone (which encircles Cape Evans) disturbance increases the *numerical* value of the vertical force, just as in the case of the northern polar region.

Kingua Fjord.

Cap Thordsen.

Fort Rae.

Point Barrow.

Jan Mayen.

Nova Zembla

Bossekop.

0h 4h 8h 12h 16h 20h 24h

0   50   100γ
VERTICAL FORCE SCALE

FIG. 13. Annual mean daily variation of the vertical force in polar regions

9.7. Disturbance-daily variation in high latitudes. The local-time (daily) additional variations due to disturbance will next be considered. Fig. 10, curves (b), (c), shows that the additional daily variation in the vertical force preserves a constant phase from the equator to as far north as Sitka, with a morning minimum and an afternoon maximum; the amplitude increases greatly with latitude. The increase persists beyond the (magnetic) latitude of Sitka—where the 'all days minus quiet days' range is about 20γ (cf. Fig. 10, curve 1c)—to Bossekop, where the range exceeds 100γ; the S_D curves (all days minus quiet days) for Z at Bossekop and other polar stations are illustrated in Fig. 13. These curves (for the stations to the north of Nova Zembla) are of similar very large range, but the most remarkable feature is that their phases are all opposite to that for Bossekop and the stations of lower latitude; the curve for Nova Zembla is transitional between the two series. The reversal of phase seems to occur within a narrow belt of magnetic latitude adjacent to the auroral zone. Such a reversal, in a region where on the two sides of the dividing line the range of the daily variation of Z is so large, constitutes perhaps the most striking of all the average characteristics of the field of world-wide magnetic disturbance.

The S_D variations (all days minus quiet days) in the *horizontal* plane are likewise of considerable magnitude in polar regions, the range being

of the order 50γ, as for the vertical force. This is far greater than in lower latitudes, and even than at Sitka, in magnetic latitude 61° (Figs. 9, 11, curves 1 c). In passing northwards from Sitka and crossing to within the auroral zone, the daily variation of the force in the horizontal plane undergoes a striking change of type, which is not simply

SITKA   SODANKYLA   BOSSEKOP

KINGUA-FJORD   CAPE EVANS
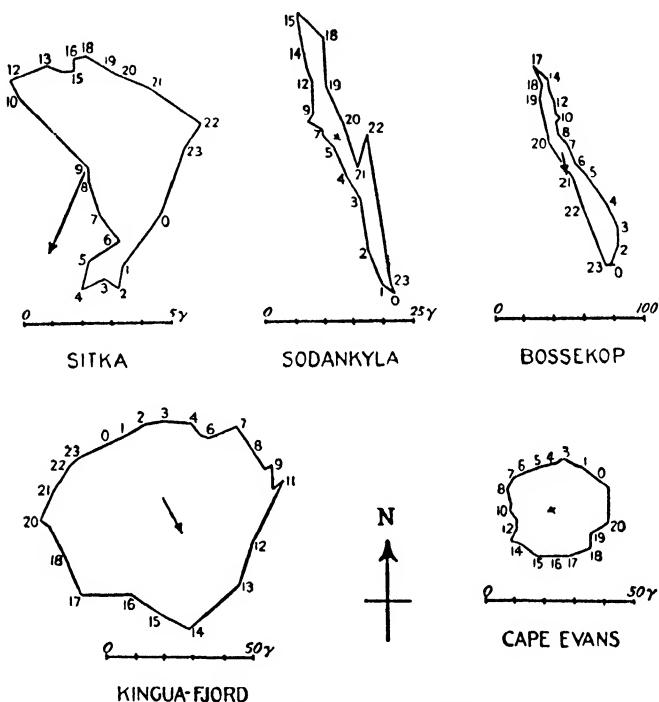
FIG. 14. Horizontal-force vector diagrams for high latitudes (north or south)

a reversal, as in the case of the vertical force. The change is best shown by the *vector* diagrams of the horizontal-force S_D variation, as in Fig. 14. In this figure the diagrams for the different stations are drawn on different scales, as indicated, according to the intensity of S_D at each station. The mean horizontal disturbance vector (§6), illustrated in Fig. 12, is shown also (to scale) in Fig. 14, in which the geographical north direction is upwards. This vector is, as Fig. 12 shows, roughly in the magnetic meridian.

At Sitka the vector diagram still bears some slight resemblance to the roughly oval form, elongated in the direction transverse to the magnetic meridian, shown at Greenwich and other stations in similar latitudes. But for stations quite near to the auroral zone, like Ssagastyr,

Sodankyla, Nova Zembla, and Bossekop, in magnetic latitudes 61°, 66°, 64°, 67°, the diagram is very narrow in this direction (i.e. parallel to the zone), and is elongated in the direction normal to the zone; the maximum poleward force occurs at about 16^h or 18^h, and the opposite minimum at about 2^h. On passing well inside the zone the vector diagram again becomes oval, indeed nearly circular; this is illustrated by the diagrams for Kingua Fjord and Cape Evans, in magnetic latitudes 78° N. and 79° S.; note that these two curves are described in opposite senses. The curves for other stations in or near the zone, like Cap Thordsen, Fort Rae, and Point Barrow, are of an intermediate and less simple type. The horizontal-force disturbance vectors illustrated in Figs. 12 and 14 show the relative magnitude of the disturbance change in the *mean value* of the force in the horizontal plane, and of the daily varying *departure* from that mean.

9.8. Summary of the average characteristics of disturbance in polar regions. The average characteristics of disturbance in polar regions, considered in §§ 6, 7, are those associated with relatively slight disturbance, but the results described in §§ 4, 5 suggest that the distribution and development of the additional disturbance field in polar regions, as well as in lower latitudes, remain fairly constant and independent of the intensity of the disturbance. There is, indeed, some polar evidence for this conclusion; this evidence will be briefly reviewed.

It has been seen that the additional disturbance-daily variations S_D in polar regions are very intense (the range being of the order of 50γ) as compared with the corresponding variations in lower latitudes. Even on the five international quiet days per month there remains a certain amount of disturbance, with its associated disturbance-daily variation. In middle and low latitudes this is small compared with the normal quiet-day variation, but it is still considerable in high latitudes, sufficiently so to be able partly to mask the residual daily variation corresponding to an ideally quiet day.

Within the auroral zone the similarity between the S variations on quiet, average, and disturbed days is notable, as was shown by Chree for the Antarctic station Cape Evans, which is inside the southern auroral zone. Fig. 15 reproduces his horizontal vector diagrams of S for quiet and disturbed days in midwinter at this station. On the assumption that at such stations S even on quiet days mainly consists of S_D, these diagrams constitute another partial confirmation of the view that the general character of the D field, in polar as well as in

lower latitudes, preserves a constant general type over a wide range of variation of its intensity.

J. M. Stagg, leader of the British expedition to Fort Rae in connexion with the second International Polar Year, carefully examined the variation of the type of S with the degree of disturbance. In his report of that expedition [G 94], he considered S both for the international quiet and disturbed days, and also for still quieter (q') and still more

x————x Disturbed days (5 a month)

o··········o Quiet   days   " "   "
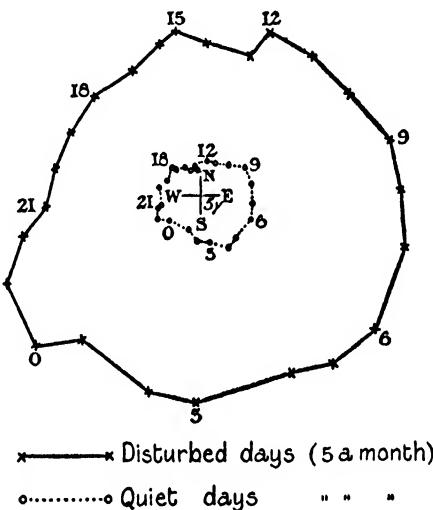
Fig. 15. Horizontal-force vector diagrams for Cape Evans (Antarctic) for quiet and disturbed days in midwinter

disturbed (d') days, numbering 38 and 40 respectively, in the thirteen months' observations. He concluded that in Z at Fort Rae, in magnetic latitude 69°, S_q is an appreciable constituent of the whole S for the very quiet q' days, and that its range is not negligible compared with that of S_q at more southerly stations; consequently in Z the S curves for q' days differ notably in type from the curves for disturbed days. But in H and D the proportion of S_D in the S curves even for the q' days seems quite considerable (G 94, p. 180); this is in accordance with Chree's horizontal vector diagrams for Cape Evans.

9.9. Evidence in S_D for a shift of the auroral zone. The conclusion that the type of the D field is fairly constant over a wide range of variation of its intensity is quite consistent with the possibility (for which there is observational evidence) that at an individual polar station near the auroral zone there may be considerable change of type in the disturbance-daily variation, or in the magnitude or sign of the disturbance

change in the *mean* magnetic force, as the intensity of disturbance increases. This is because of the close relationship of the average disturbance field in polar regions with the auroral zone. This zone appears to broaden and move towards lower latitudes during periods of intense disturbance, so that a station which ordinarily is on the equatorial side of the zone may during a magnetic storm be under, or on the polar side of, the zone. If so, the disturbance-changes in the magnetic field at the station may be radically different during slight disturbance from those during magnetic storms, even though, in *relation to the zone*, the general character of the disturbance field is similar in the two cases.

The station Nova Zembla seems to exemplify this possibility. The average Z variation for this observatory (Fig. 13) is of transitional type, suggesting that the observatory is nearly under the dividing line between the two types of Z variation referred to in Fig. 12 (this dividing line is probably identical with the auroral zone). But on many individual days the daily Z variation is of definite type, sometimes with morning maximum and afternoon minimum, and at other times with the reverse phase. Out of about a year's observations, 100 days of morning Z maximum and about 112 days of afternoon maximum were found. On the remaining days either the record was faulty or the maximum and minimum Z both fell in the forenoon or afternoon, with less than six hours between; most of the very highly disturbed days were in this class. The mean range of the horizontal-force variation on the above two classes of days was determined, and was found to be distinctly greater on the days of morning than of afternoon Z maximum (429γ as against 268γ). This suggests that on the more disturbed days Nova Zembla may be inside the auroral zone, while on quieter days it is outside it. It would be of interest to examine the records of other observatories, near the auroral zone, in a similar way.

9.10. The average disturbing vector during storms. The discussion in §§ 5, 6 corresponds in part to an investigation by van Bemmelen [5] of what he called the 'post-perturbation' or after-disturbance effect (in German, *Nachstörung*). He systematically studied the changes, from day to day, in the daily means of the three magnetic elements at many observatories, during many days following upon a number of magnetic storms. Immediately after a storm the horizontal force is generally below the average, but it increases during the following days. At many stations, especially in high latitudes, the declination is also affected, while the vertical force 'mostly decreases, though not with the same regularity'.

The steady increase of horizontal intensity during quiet periods may be illustrated by Fig. 16, due to McNish [7a], for Watheroo (West Australia). The long continuance of the increase, when the conditions are suitably quiet, has also been strikingly shown by Schmidt for Potsdam (Fig. 17).
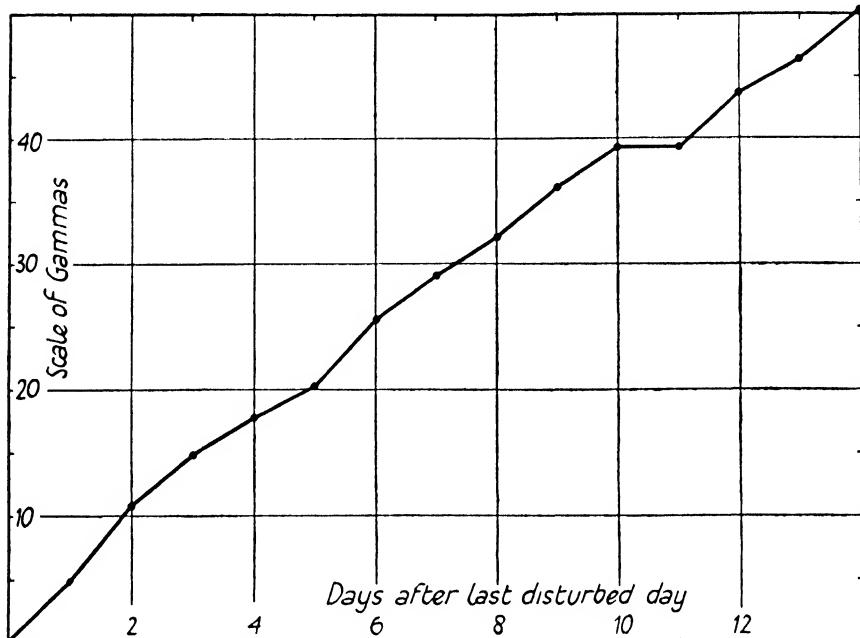
FIG. 16. The increase of horizontal intensity at Watheroo magnetic observatory on quiet days in 1919, 1925–9, derived from successive daily means. (After McNish)
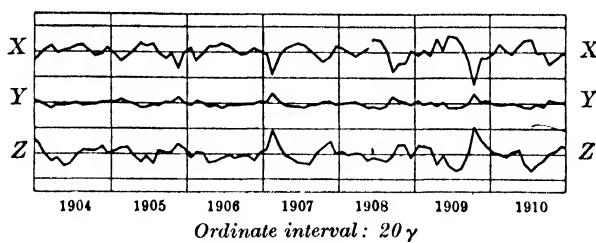
FIG. 17. Monthly means of the force components at Potsdam, 1904–10, expressed as deviations from the annual means. The horizontal lines are drawn 20 γ apart. (After Schmidt)

van Bemmelen found that at each station the mean change of the magnetic force in the horizontal plane, from day to day, has a characteristic azimuth; this is naturally not shown in each individual series of days after a magnetic storm, but appears clearly in the mean of several series, whether or not the days after the storms continue to be

disturbed or are quiet. The changes in consecutive periods of a day decrease (on the average) in a geometrical progression, corresponding to an exponential decay of the storm-time field.

The geographical azimuth of the disturbing horizontal force at twenty-six stations was found by van Bemmelen to be as follows.

Azimuth of the horizontal disturbing force at twenty-six stations

The azimuth is stated in degrees, to the east or west of geographical south.

Godthaab . . . . 28 E.	Tiflis . . . . 15 E.
Kingua-Fjord . . . . 27 E.	Nova Zembla . . . 22 E.
Point Barrow . . . . 24 W.	Barnaul . . . . 11 E.
Fort Rae . . . . 4 E.	Nerchinsk . . . 9 W.
Toronto . . . . . 1 E.	Zikawei . . . . 6 W.
Washington . . . . 7 E.	Manila . . . . 3 E.
Cap Thordsen . . 40 E.	Bombay . . . . 8 E.
Jan Mayen . . . . 42 E.	Batavia 0
Bossekop . . . . 25 E.	St. Helena . . . 12 W.
Sodankyla . . . . 18 E.	Cape of Good Hope . 13 W.
St. Petersburg—Pavlovsk . . 20 E.	South Georgia . . . 8 E.
Greenwich . . . . 18 E.	Cape Horn . . . 12 E.
Pola . . . . . 8 E.	Hobarton . . . 12 W.

On setting out these directions on a map, van Bemmelen found that they were very nearly perpendicular to the isochasms, or lines of equal frequency of the aurora polaris; thus the horizontal disturbing forces appear to radiate from a point situated between the geographic and magnetic north poles, approximately coinciding with the pole of the earth's magnetic axis. This has been confirmed by Slaucitajs and McNish [7], who considered the horizontal vectors at many stations, representing the difference between the daily means of X and Y (international disturbed minus international quiet days) for 1927; their results are given in Fig. 18, which shows also some of the geomagnetic meridians (p. 105).

Fig. 19 is a stereoscopic diagram [2.79] showing the whole disturbing vector (international disturbed minus quiet-day means, taking account of all three components X, Y, Z) at Potsdam. The vector lies nearly in the vertical plane through the geomagnetic meridian, and is intermediate between the horizontal direction and the direction of the earth's magnetic axis.

van Bemmelen also considered the daily variation of the disturbing force at the four observatories Pavlovsk, Tiflis, Batavia, and Cape Horn; he found this to be of the same type whether the days following the storm were still disturbed or not. This daily inequality corresponds to a combination of the storm-time and local-time variations described
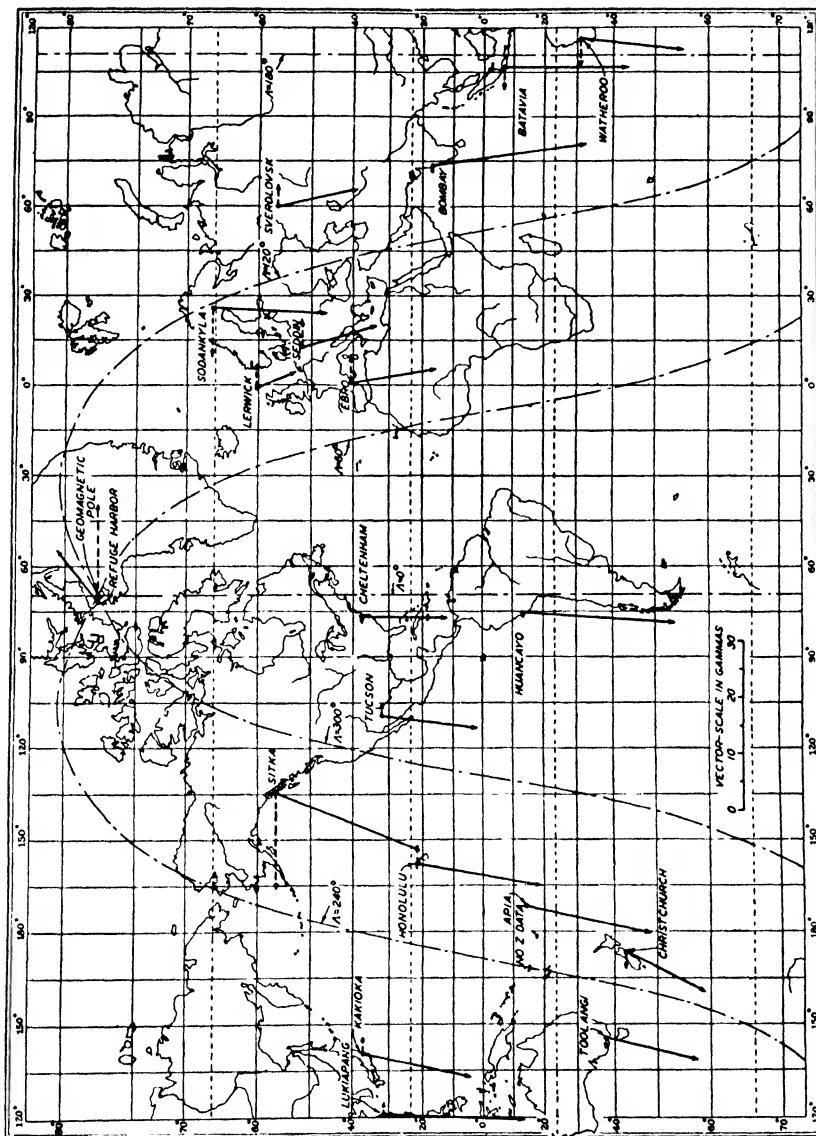
Fig. 18. Mean vector differences for disturbed minus quiet days, 1927. Full line: horizontal intensity; broken line: vertical intensity (positive towards left). (After Slaucitajs and McNish)

in §§ 2, 3. Mention may also be made of a discussion by Lüdeling [9] of the disturbing force at polar stations, using the data for the polar year 1882–3
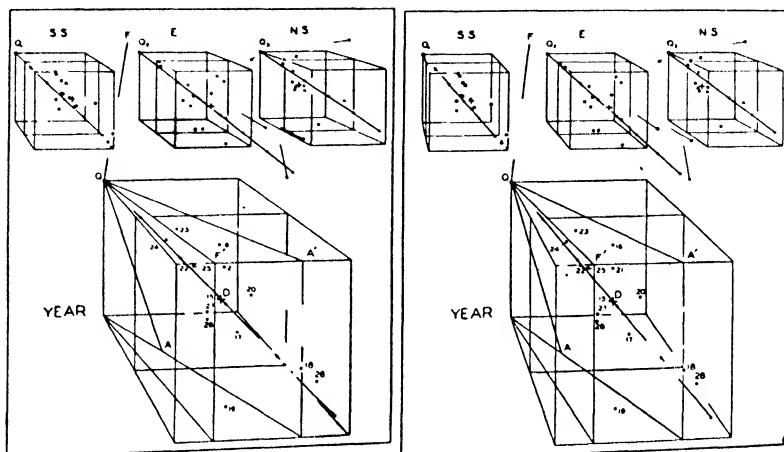
FIG. 19. Stereogram showing the 'disturbed-day minus quiet-day' differences for the mean magnetic vector, at Potsdam, for each of the years 1915–28. The edge of the cube, orientated astronomically, represents 10γ. The point Q in the far upper left-hand corner represents the origin of the difference-vectors. Lower part: annual means; upper part: means for the three seasons. QA is parallel to the earth's magnetic axis. (For full description see [2.79])

9.11. The initial impulse in sudden commencements.

Another important investigation by van Bemmelen [10] related to the starting impulse, or sudden initial change in the magnetic elements, that very often accompanies the outbreak of a storm. Fig. 37 (p. 336), due to W. J. Peters [11], shows for several stations the H records (all to the same scale) of the sudden commencement of the great magnetic storm of 1921 May 13 (p. 336).

For Batavia van Bemmelen examined 131 cases occurring between 1882 and 1899. In *all* cases ΔH (i.e. the change in H) was positive; in nearly all cases ΔD was westwards, though in 12 of them there was a slight initial eastward movement; in all but 6 of the cases ΔV (or $-\Delta Z$) the change of vertical force, which at Batavia is upwards, was negative, and in the few positive cases the change was relatively slight. In H and Z the initial rapid increase or decrease continues, but in D it stops or is inverted. The duration of the rapid initial change, in the mean of 124 cases, was 4·5 minutes for H, 3·2 minutes for D, and 12·0 minutes for Z. The following list shows the distribution of the durations in the case of ΔH, and the corresponding mean values of ΔH; it is clear that

ΔH is nearly independent of the duration, so that the rate of change of H is inversely proportional to the duration of the change.

Duration	Number of cases	ΔH (γ)
0–2 min.	15	53
2–4 ,,	45	43
4–6 ,,	35	42
6–8 ,,	20	33
8–15 ,,	6	46

The average values of the movement in the four different quarters of the day are as follows:

	ΔH	ΔD	$\Delta V (= -\Delta Z)$
0–6^h	45 γ	7' W.	$-11\ \gamma$
6–12^h	41 γ	10' W.	$-16\ \gamma$
12–18^h	52 γ	7' W.	$-16\ \gamma$
18–24^h	40 γ	8' W.	$-11\ \gamma$

The average direction of the disturbing vector in the horizontal plane, determined from 29 cases recorded between May 1906 and November 1907 at Buitenzorg (near Batavia), was 21° W. of N.; twenty of the individual values ranged between N. and 58° W. of N. A curious difference between Batavia and Buitenzorg was found, in that at the latter the (upward) vertical force first increased, and then decreased in the way observed at Batavia; the introductory movement at Buitenzorg preceded that at Batavia: it started about 0·3 minute after the ΔH movement began, and lasted from 1 to 3 minutes before the vertical force (at both stations) began to decrease.

Data from other observatories were collected for 17 of the cases (1892–1905); the observatories included Greenwich, Paris, Perpignan, Coimbra, Zikawei, San Antonio (Texas), Honolulu, Bombay, Manila, Samoa, Mauritius, and Melbourne, ranging in latitude from 51° N. to 38° S. It proved difficult to be certain that the movements in the three elements, and at different stations, were simultaneous to within 1 minute.

In almost all cases at all the stations ΔH was positive (northwards), though in a few cases there was a minor introductory negative movement, while in one or two others the positive movement was minor and introductory to a larger negative one.

The sign of ΔZ on different occasions at the same station was not so constant as for ΔH, but at each station there was a predominant sign. This was not the same at all stations: at Greenwich, San Antonio,

Manila, Samoa, and Mauritius it was positive, while at Paris, Zikawei, and Melbourne it was negative. The opposition of sign between the values of ΔZ at the relatively near stations of Greenwich and Paris (Parc St. Maur) was striking and almost invariable, though in ΔH and ΔD the movements were generally the same. The opposition of ΔZ
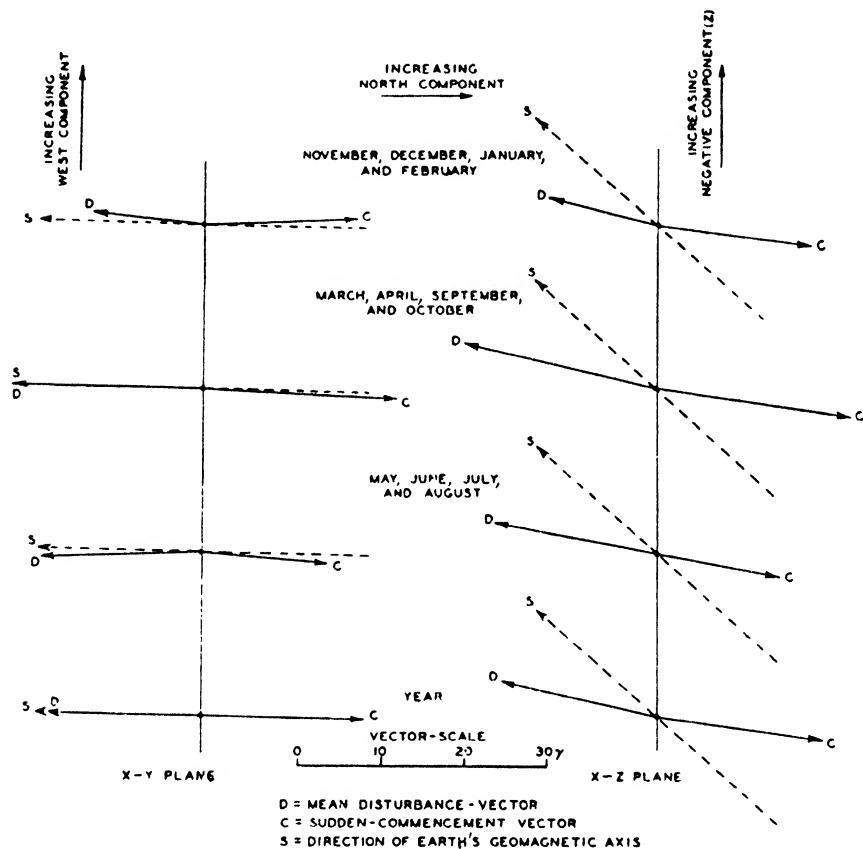
FIG. 20. Mean disturbance vectors and sudden-commencement vectors at Watheroo magnetic observatory, 1919–30. (After McNish)

appears only to apply to sudden changes, whether occurring at the beginning of a storm or during a disturbance that has been in progress for some time; during the slower oscillations which follow, the movements at the two observatories are similar.

The sign of ΔD on different occasions at the same observatory was more variable than for ΔH or ΔZ. But, unlike ΔZ, the sign of ΔD was generally the same, for each disturbance, at relatively near stations. At more distant stations, however, corresponding values of ΔD are

often opposite. In particular, ΔD was often of opposite sign in Europe and North America.

Peters [12] has constructed stereograms showing the force vectors of the sudden commencements for three storms. They lie mostly along or nearly along the earth's surface.

McNish [13], from 151 sudden commencements at Watheroo, 1919 to 1930, found that the average vector of the sudden commencement is directed almost opposite to the main disturbance vector, that is, to the deviation of the force vector, from the normal vector, at the time of strongest disturbance. Both vectors lie approximately in the plane through the magnetic axis and the station, but are definitely less inclined to the earth's surface than is a parallel to that axis. This is illustrated by Fig. 20, which shows the projections of the various vectors on the horizontal (X, Y) plane, and on the vertical (X, Z) plane. As regards the main disturbance vector, it confirms for Watheroo the result for Potsdam shown stereoscopically in Fig. 19.

The intensity of sudden commencements seems to depend to a certain extent on local time; that is, although sudden commencements of great storms are world-wide, the initial movement is more or less strongly developed at different places, so that it may remain unnoticed at certain stations.

Since the discovery of the magnetic effects accompanying radio fade-outs over the daylight hemisphere (10.3), the older statistics of sudden commencements may have to be revised because of the (superficial) resemblance of these two magnetic phenomena.

9.12. Symmetrical-zonal and asymmetrical features of disturbance. The preceding account summarizes what appear to be the principal *average* characteristics of the field of world-wide magnetic disturbance. It is clear that the field is not only more intense but also more complex in structure in the polar regions than elsewhere.

The complications of the disturbance field in high latitudes may be divided into two classes, one of which is associated with the existence of a band or zone (the auroral zone)—at ordinary times comparatively narrow—across which there is a complete change of type of the disturbance-daily variation, and also of the Z storm-time variation. This part of the polar complexity of structure of the disturbance field comprises what may be called the symmetrical-zonal features of the field, that is, the features connected with the existence of the zone, but abstracting those which depend on the asymmetry of the zone relative to the geographical axis of the earth. The further complications due

to the eccentricity of the zone may be called the asymmetric features of the field; these correspond to the differences between the variations, with respect to local time, at different stations along the same parallel of latitude. The angular distance between the north pole and the poles of the dipole axis (or the centre of the auroral zone) is comparable with the angular radius of the auroral zone, and the divergence must and does produce important consequences for the disturbance field in the polar regions, where the field is so highly differentiated; in low latitudes, on the other hand, the asymmetric features of the disturbance field are relatively unimportant.

The geographically asymmetrical characteristics of the D field have been studied by Vestine [39]; he finds that the auroral-zone curve indicated by the reversal of S_D in Z (or the *geomagnetic auroral zone*) agrees roughly with the zone of maximum auroral frequency given by Fritz, except in regions where his auroral data were scanty; it appears to undergo regionally a small daily oscillation. With respect to this geomagnetic auroral zone, S_D in high latitudes appears to have a simple distribution, mainly dependent on geomagnetic latitude and geomagnetic local time. This suggests that the auroral zone has a controlling and initiating influence over the world-wide electric currents responsible for S_D. In effect, the D field has a simple character relative to the geomagnetic axis, and its geographical asymmetries are almost entirely due to the obliquity of the geomagnetic to the geographical axis.

9.13. A current-system corresponding to the D field. As in Chapter VII for S_q, so here for the D field, atmospheric current-systems will be described that could give rise to a magnetic field, at the earth's surface, having the most important symmetric-zonal features of the observed disturbance field. The current-system affords the simplest means of representing these features of the field, and may be regarded solely as such a means; there is no implication that the field actually is produced by such a set of currents flowing parallel to the earth in the atmosphere; it may be noted, however, that the field could *not* be produced mainly by a set of currents flowing *within* the earth; such currents are responsible for only a minor part of the field, a part presumably induced by the primary outer current-system (Ch. XXII).

The current-systems to be described [19] relate to an ideal earth having coincident magnetic and geographic axes. In making comparisons with the actual earth, a station on the ideal earth, in a given situation relative to the ideal symmetric-auroral zone, must be compared with a station on the actual earth in a similar situation relative

to the actual auroral zone. The hypothetical and the actual disturbance fields will, of course, not agree in all particulars, but the ideal current-system, and its hypothetical field, will be of value according to the extent to which they reproduce the observed features of the disturbance field; and a study of the differences between the hypothetical and observed variations is of value in the determination and explanation of the asymmetric features of the field, as has been shown by Vestine [39].

The current-systems illustrated below in Figs. 21–26 refer to magnetic storms of moderate intensity, being, in fact, based on the results given in Figs. 4, 6, 7, 8. The maximum storm-time reduction of H at the equator, in the 'average' storm of this type, is 40γ. The current-systems are drawn for the epoch of this maximum intensity of the storm; at other epochs the current-distribution may be fairly similar in form but of lower intensity, except that during the first phase, when H is increased, the current-direction must be reversed. Whether the system at that time is of the same type as at maximum phase is not yet known.

As in Figs. 7.15, 7.16, the current-diagrams of Figs. 21–26 show both the direction of current-flow and also the current-intensity, being drawn at intervals such that between adjacent lines the total current-flow is 10,000 amperes. In the auroral zones the flow is so intense that in the diagrams it has not been possible in all cases to indicate the individual current-lines, which appear merged in a dense mass.

In the foregoing discussion the disturbance magnetic field has been analysed into three parts, the daily or local-time part S_D, the storm-time part D_{st}, and the irregular part D_i. The current-system responsible for the whole field can be similarly analysed, since the magnetic fields of superposed current-systems are superposable.

9.14. The S_D current-system. The ideal current-system proposed as corresponding to the *daily* part of the disturbance field, S_D, is indicated in Figs. 21, 22. They show the atmospheric current-system as seen in 'elevation' from the sun, and also in 'plan' as seen by a distant observer looking towards the north pole from a point far above the pole, along the earth's axis. The ideal auroral zones, or zones of great current-intensity nearly coinciding with them, are indicated by the crowded current-lines; in Fig. 21 the wedge-shaped appearance of the auroral current is not to be interpreted as a feature of Nature; it is due only to the greater intensity of the current in the auroral zones, at 6^h and 18^h local time, which in the diagram is indicated by many
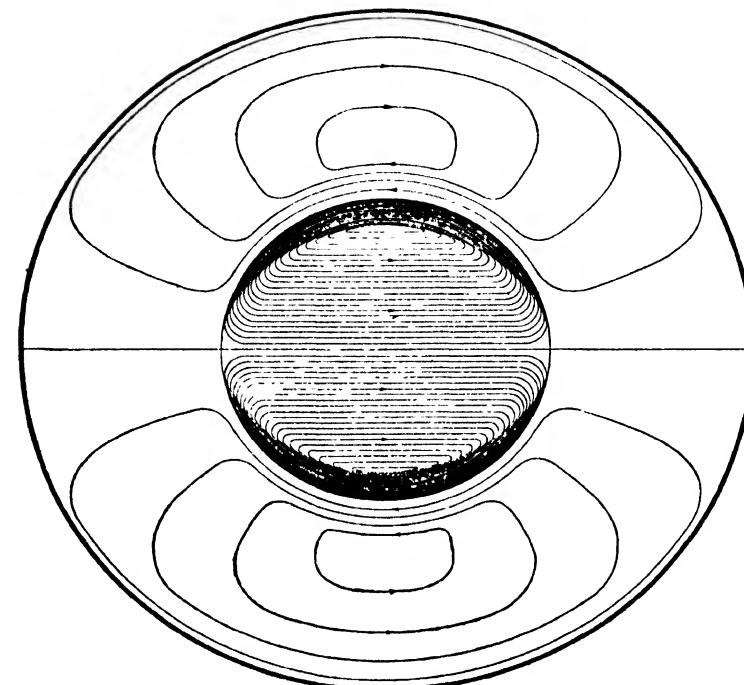
Fig. 22. View from above the north pole
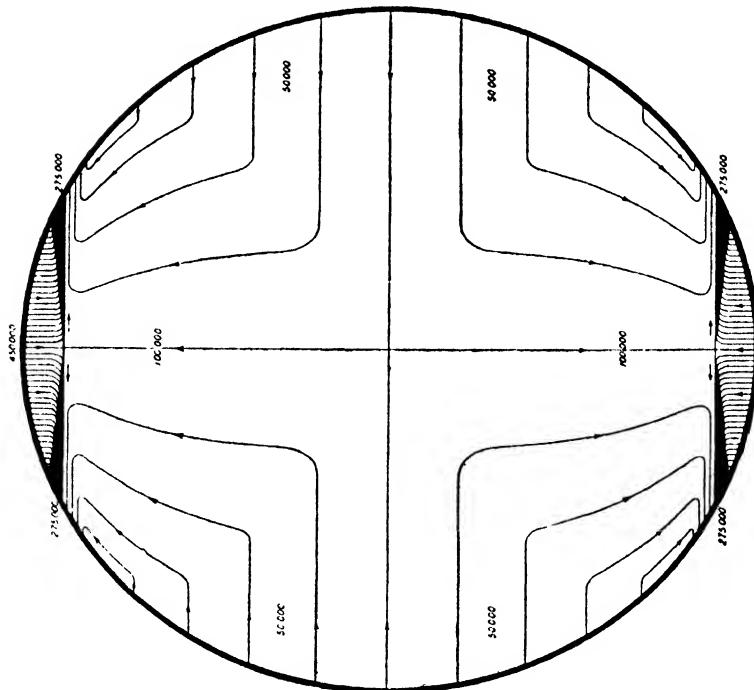
Fig. 21. View from the sun

Views of the idealized overhead electric current-system that could produce the S_D field

current-lines crowded together. The currents are specially intense not only along the auroral zones but also over the polar caps within these zones, as is indicated by the greater intensity of the disturbance field in polar regions. Most of the currents flowing along the auroral zone (in opposite directions in the two halves of each zone) are supposed to complete their paths by flowing across the polar caps within the zone; but a small part of the zonal currents may complete its circuits over the middle belt of the earth. The intensity distribution of the current-system is inferred from the data of Figs. 6–8, in a manner similar to that described, for S_q, in 7.6; it makes allowance for a fraction of the observed field being produced by currents induced within the earth. The much greater magnitude of the *zonal* and *polar cap* currents than of those between the north and south auroral zones is indicated by a comparison of the total flow across one polar cap (indicated schematically in Fig. 22 by parallel lines), which is 450,000 amperes, with the flow in any one of the four inter-zonal circuits, namely 50,000 amperes. The general intensity of the *inter-zonal* part of the S_D current-system during a moderate magnetic storm is thus comparable with the intensity of the S_q current-system (Figs. 7.15, 7.16); on average days (as indicated by 'all-day minus quiet-day' inequalities) the S_D current-system is only about one-fifth as intense as that shown in Figs. 21, 22, so that the part of it between the zones is decidedly less important than S_q on such days—and on quiet days it is still smaller. The part of the S_D system over the polar cap is more intense than the S_q system even on the average day, and far more intense during storms.

The maximum S_D current along the auroral zone, eastwards at 18^h and westwards at 6^h, is shown as 275,000 amperes, of which only 50,000 are due to the inter-zonal current-circuits.

The space distribution of the current-systems of Figs. 21, 22 is supposed to be constant as viewed from the sun, just as for the S_q current-systems of 7.6; the earth revolves eastwards relative to the current-system, or the current-system westwards relative to the earth; in the atmosphere above any particular locality on the earth, the currents are continuously modified, to the magnitude and direction characteristic of its situation at the time, relative to the meridian through the sun; this situation is determined, of course, by the latitude and local time of the station. The magnetic field of the current-system is observed as a field which varies throughout the day at each station, according to local time.

It can easily be verified that Figs. 21, 22 do correspond to the nature

of the S_D daily variations in the three components of magnetic force, at a point P at ground-level—that is, below the current-system.

At each such point P, the horizontal component of the S_D field may be assumed to be due mainly to the current in the region overhead and adjacent; this current will produce a field, at points immediately below it, whose horizontal component is not much different from that of an infinite plane horizontal current-sheet, in which the current-flow everywhere has an intensity and direction equal to that of the actual current-system above P. The field of such a sheet is horizontal, the direction, at points below the sheet, being displaced 90° from that of the current, anti-clockwise as viewed from above. For example, this implies that in Fig. 21, along the equator, on the right half of the figure (corresponding to the hours 12 to 18 of local time), the horizontal force due to the current-sheet, at points on the ground, is southwards, whereas on the left-hand half (hours 6 to 12) it is northwards. In the centre of Fig. 21, where the current-direction turns sharply, its intensity must be zero, and there the horizontal-force vector due to the current-system must likewise be zero. Along the equator the current-intensity is greatest, in the figure, at 6^h and 18^h, and there the S_D variation of horizontal force will have its respective maxima and minima. Similarly, on both sides of the equator, to the latitudes (N. or S.) at which the current-circuits between the two zones shrink to points, the horizontal north force will have its maximum at 6^h and its minimum at 18^h local time; above these latitudes, up to the auroral zones, the north force variation will be reversed. This agrees with the general character of the curves (b, c) in Fig. 6. The variation of west force or declination will be of opposite phase on the two sides of the equator, without further reversal up to the auroral zones; the west force in northern latitudes, up to the zones, will have its maximum at 12^h and its minimum at midnight. These inferences from Figs. 21 and 22 are in general agreement with the curves b, c of Fig. 8.

It is not quite so easy to infer, from current-diagrams such as those of Figs. 22, 23, what will be the vertical component of magnetic force due to the currents; on the two sides of any point P' in the current-sheet, immediately above a point P on the ground, the horizontal current-flow will generally have approximately the same direction, and will make opposite contributions to the Z force at P. The resultant sign, and some rough idea of the relative magnitude of the Z force, can be inferred by considering the probable intensities, and the extent, of the currents on the two sides of P'. In Fig. 21, for example, all along

the equator, and also on the 0^h and noon meridians, it is clear from the symmetry of the current-sheet that there it produces no Z-component. At the northern foci of the inter-zonal current-circuits, at about 55° northern latitude on the meridians 6^h and 18^h, Z will be upwards at 6^h and downwards at 18^h; near the zone it will be specially intense at these hours, and will have the same sign, just south of the northern zone, as at all lower northern latitudes, whereas just within the zone the signs will be reversed.

In any latitude between the equator and the northern auroral zone, the daily maximum of downward force will occur at 18^h, and the daily minimum at 6^h. The daily range will increase from zero at the equator to a maximum at a latitude only slightly less than that of the centre line of the auroral zone. These inferences are in general agreement with the observed facts summarized in the curves b, c of Fig. 7.

The range of the vertical-force variation will decrease rapidly from the latitude of maximum (just south of the zone) to zero at stations directly under the zone. Farther north still, the variation will be reversed, and its range will rapidly rise to another maximum a little north of the zone, and from there it will gradually decrease to zero at the pole or centre of the zone.

These inferences are in general agreement with the facts summarized in Fig. 13, and an oval curve can be drawn on the map of the Arctic region (Fig. 12) so as to separate the stations having vertical-force variations of opposite phase, in such a way that those stations within the curve have maximum downward vertical force in the forenoon, while at those outside the curve it occurs in the afternoon. The curve must be drawn (cf. Fig. 12) so as to pass to the north of Fort Rae and Bossekop, and nearly over Nova Zembla. It will therefore not agree exactly with the zone of maximum auroral frequency as drawn by Fritz, though its departure from the latter is not great. The magnetic data are unfortunately still insufficient to indicate the complete course of the curve referred to. Moreover, too great reliance cannot be placed on the accuracy of the auroral zone as determined by Fritz (p. 469).

As regards the daily variations of magnetic force in the horizontal plane in the polar regions, the observed variations of the two components, north and west, have already been shown to accord with Figs. 21 and 22 for stations up to the latitude of Sitka. As the current-zones are approached, the horizontal-force variation will become more and more directly transverse to the zone; the west-component variation will decrease in range till it vanishes under the zone, whereas the north-

component variation will remain the same in phase (morning minimum and afternoon maximum) and increase greatly in magnitude. The vector diagram of the horizontal-force variation will thus become rectilinear, with its direction normal to the zone, and with its north and south elongations at 18^h and 6^h respectively.

On crossing the zone, the north-component variation will preserve its phase, but its range will begin to decrease; the west-component variation will reappear with reversed phase. The horizontal-force vector diagram will therefore again become oval, but will be described in the counter-clockwise direction instead of in the clockwise direction (as is the case just south of the northern auroral zone); the horizontal-force vector at stations just within the zone will be southerly at 6^h, easterly at 12^h, and so on.

Still farther within the zone, the north-component variation will vanish and reappear with reversed phase; the west-component variation will have the same phase everywhere within the zone. The direction of description of the vector diagram will thus be again reversed; this is the third reversal in the direction of description of this diagram, in proceeding northwards from the equator (the three reversals occur at about the latitudes 55°, 65°, and 70° or 75°, the middle latitude being that of the zone itself): the direction of description in the central region within the zone will therefore be clockwise, the force vector being northerly at 6^h, easterly at 12^h, and so on. A simple way of regarding this variation, in the vicinity of the pole, is to consider it as due to the rotation of the earth within a stationary magnetic field, which near the poles is nearly uniform and horizontal: relative to the earth, therefore, the variation at any particular station near the pole will appear as the uniform rotation of a nearly constant vector, in the direction opposite to the rotation of the earth. The constant vector, according to Figs. 21 and 22, will be in the meridian plane normal to the radius vector to the sun, or, otherwise stated, in the meridian plane of the 6^h and 18^h meridians of local time. This gives rise to north- and west-component variations with phases as just described: also the vector diagram will be nearly circular. The reference here is to the Arctic cap; in the Antarctic the horizontal component of the force near the pole will be opposite in direction to that at the south pole, as Fig. 22 indicates; and the rotation of the earth will cause the vector diagram at stations near the south pole to be described in the anti-clockwise direction, the vector being northerly at 6^h, westerly at 12^h, and so on.

The hypothetical horizontal-force daily changes, or vector diagrams,

in the polar regions, thus deduced from Figs. 21 and 22, are mostly in fair accord with observation, though there are some discrepancies. The transition from the vector diagram at Sitka to the rectilinear vector diagram at Bossekop, which is under or very near to the auroral zone, is in accord with this discussion, though the observed phases do not always agree with the hypothetical phases. In the diagrams for Bosse-kop and Sodankyla, for example, the extreme elongations are at 0^h and at 15^h or 17^h, instead of at 6^h and 18^h, as the hypothetical current-diagrams suggest: the diagram for Ssagastyr (not included in Fig. 18) has its extreme elongations at 4^h and at 18^h or 20^h. Thus at different stations near the auroral zone the maximum current along the zone seems to occur at different hours of local time: this is an example of the consequences which may be attributable to the asymmetry of the zone about the geographical axis. In high northern and southern magnetic latitudes, such as at Kingua Fjord and Cape Evans, the observed vector diagrams are nearly circular, and are described nearly uniformly, in the directions inferred from Figs. 21 and 22, but again there are phase-differences of about 3 hours between the actual and the hypothetical diagrams. Such differences are not surprising in view of the very considerable asymmetry of the auroral zones with respect to the geographical poles. But it may reasonably be claimed that the relatively simple current-system of Figs. 21 and 22 suffices to account for a large proportion of the multitude of facts summarized in Figs. 6–8.

9.15. The storm-time current-system. The second part of the idealized current-system will next be considered, that, namely, which is responsible for the storm-time part of the varying disturbance field. By virtue of its definition and mode of derivation, this part is essentially symmetrical about the earth's axis; its intensity and its distribution in latitude vary materially during the course of a storm, as is shown by the storm-time curves given in Fig. 4 or Fig. 5 (for middle and low latitudes only, up to about 60°). Since it does not seem possible as yet to deduce similar curves for the polar regions, it is necessary to base the calculation of the relative current-intensity, in different parts of the system, on the time-average of the storm-time changes as derived from the difference between the daily means of all days and quiet days; afterwards we multiply the intensities by an estimated factor (taken as 5) to make them refer to the epoch of maximum disturbance during a moderate magnetic storm. The current-diagram thus obtained is given in Figs. 23 and 24, which refer to the epoch of maximum intensity for the storm-time part of the disturbance field. It will therefore not

Fig. 24. View from above the north pole
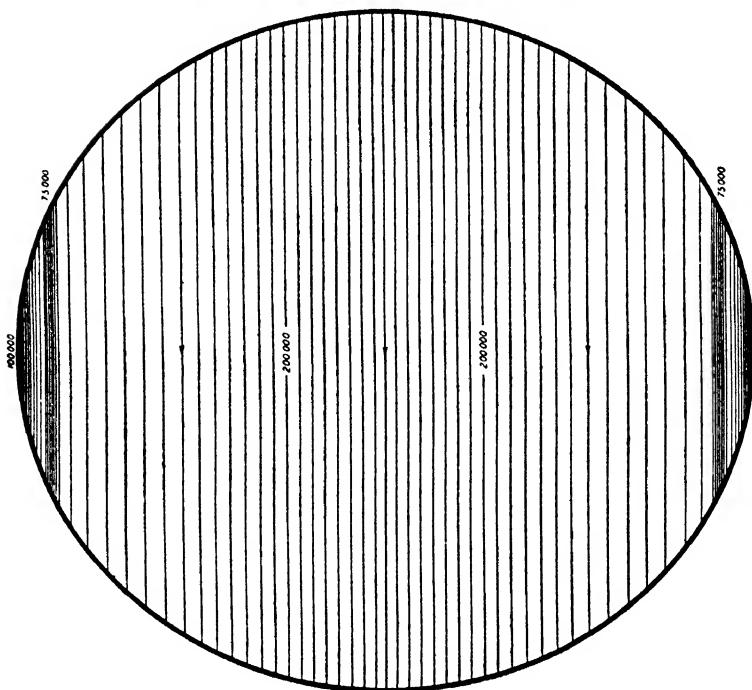
Fig. 23. View from the sun

Views of the idealized overhead electric current-system that could produce the field of the average storm-time disturbance

represent the current-system of this part of the field during the early hours of a large disturbance, when the storm-time changes have the opposite sign (in middle and low latitudes, at least).

The averaged storm-time changes have been described in § 2; they constitute a more limited set of facts than those which Figs. 21 and 22 were designed to explain. The storm-time current-system consists of currents which are everywhere in the same direction, along circles of latitude from east to west; they are concentrated with greater intensity than elsewhere, along the auroral zones, as in Figs. 21 and 22. The associated magnetic field will everywhere be directed southwards, and will be most intense in and near the current zones; this agrees with the diminution of north force observed everywhere in middle and low latitudes, and with the diminution, near the auroral zone, of the component of horizontal force normal to the zones. The vertical force due to these currents will clearly vanish at the equator, and be large, and of opposite signs, adjacent to and on opposite sides of the zones. Just south of the northern zone it will be upwards, and just north it will be downwards. These indications are in general accord with the facts, and have been confirmed by Vestine and Chapman by examination of the vertical force data obtained during the second International Polar Year, 1932–3 [20 and 39].

The total estimated westward *inter-zonal* current-flow, between the two auroral zones, in Figs. 23 and 24, is 400,000 amperes, which is twice as great as the combined currents of the four S_D circuits (Figs. 21 and 22) which occupy the same region. The estimated *zonal* current-flow along either auroral zone is 75,000 amperes, and *within* each *polar cap*, up to the pole, 100,000 amperes. These estimates are rather uncertain, but they are probably right in attributing a greater value to the total current (though not to the current-*intensity*) in the belt between the zones than to that in the 'caps' within the zones.

9.16. The combined D current-system. The complete system of atmospheric currents corresponding to the D field is the combination of the two current-systems illustrated in Figs. 21–4. It is shown in Figs. 25 and 26. It indicates that in the forenoon hemisphere (0^h to 12^h local time) the current between the zones is everywhere westward, whereas in the afternoon hemisphere the current, though mainly westward, is reversed at a latitude slightly south of the (northern) zone, and is eastward from there up to the zone. Along the zone itself, the current is westward over the greater part, extending right across the forenoon hemisphere and partly into the opposite hemisphere; the maximum westward intensity, 350,000 amperes at 6^h, considerably

FIG. 26.   *View from above the north pole*

FIG. 25.   *View from the sun*

Views of the idealized overhead electric current system that could produce the 'regular' field of a geomagnetic storm: this current-system is a combination of the two illustrated in Figs. 21–4

exceeds the maximum eastward intensity, 200,000 amperes, at 18^h. On the latter meridian the current direction is reversed just within the zone; it is in this region of reversal that the strongest Z disturbance will occur.

Vestine [39] has examined the overhead polar current-systems that would correspond to a number of disturbances observed in high latitudes during 1932-3, and finds that their general form resembles that shown in Fig. 26; but the polar current-circuits are often less symmetrical than in Fig. 26, and are distorted and also rotated through an angle, about the centre of this diagram (which corresponds to the centre of the auroral zone, not to the geographical pole); the westward current often seems to attain its maximum intensity at 2^h or 3^h instead of at 6^h.

9.17. Some details of individual magnetic storms. Schmidt's vortices. The preceding discussion deals mainly with the average characteristics of magnetic storms. It is also of importance to consider disturbances individually, in order to determine the characteristics of the irregular movements, which are eliminated when many storms are superposed and averaged.

The early workers in this field included Gauss [G 3], Adams [1], Schmidt [24], and Birkeland [G 97 and 97 a]. Gauss published plates showing simultaneous records for various observatories which collaborated, from 1835 onwards, in the *Göttinger Verein*. Adams [1] collected magnetograms for a number of disturbances from many observatories for intercomparison. Fig. 1 (p. 273) is a modified version of a diagram given by him; it shows some of the records for a storm commencing on 1885 June 24 at 10.32 p.m., G.M.T.; the records were retraced with the same time-base, and grouped so as to bring out prominently the common features from widely distant stations. Unfortunately the curves were not drawn with the same scale of force; the force scales are shown in Fig. 1 by vertical lines of length equivalent to 100γ. The epochs of local noon and midnight are indicated on each curve by the letters N and M. It is clear that while some of the magnetic changes affected all the observatories, though not equally, others were different at different observatories. Adams proposed to apply the Gaussian method of analysis to magnetic storms, but did not carry out this intention.

Schmidt [24] pointed out that if we compare the magnetic curves for two observatories not very far apart, during a storm, a great resemblance is usually apparent. Closer scrutiny shows that this consists

mainly in the number and position of the single waves on the curves, while the sizes of corresponding extreme departures show important differences. There may also be differences in the times of occurrence of these departures, sometimes so much as to make maxima at one station correspond with minima at another. The farther apart the stations, the greater and more numerous are the differences, so that at widely separated points only a few corresponding features may be found.

Especially characteristic is the frequent alternation in the relations between the changes at different stations; striking similarity is soon followed by complete difference or opposition. 'This points unmistakably to *local* processes as the immediate cause of these phenomena—processes of fluctuating strength and extension, which occur now here, now there, or, it may be, simultaneously at different places, influencing the whole earth to some degree, but principally only in a more or less limited region.'

A further striking peculiarity of the storm curves is that these centres of action usually change their location continuously and irregularly. The curves are almost always made up of simple elements—usually single waves, of a fairly regular type. The D and H changes often differ in phase, which indicates a changing direction of the disturbing force, and suggests that the centre of action is in motion. Schmidt concluded that the cause of the storms consists of electric currents in the upper air, 'wandering vortices' of electricity, comparable in form and motion to the cyclones and anticyclones of the lower air.

To gain an insight into their nature, he considered the distribution of the disturbing-force vectors, and showed that those for different observatories, during strong and rapid perturbations, roughly diverge or converge from a centre, while at times of comparative quiet they are nearly parallel, indicating a distant source of disturbance. In the former case, the points of convergence or divergence are sometimes stationary, and at others move with a velocity of the order of 1 km./sec. or 60 km./min. Comparison of the horizontal and vertical components confirms the supposition that the currents are situated mainly in the atmosphere, not within the earth. These currents, however, must induce earth currents, which will contribute appreciably to the resultant disturbance, and which are actually observed.

Fig. 27 shows some typical examples, relating to a disturbance occurring on 1896 February 28, 18^h–19^h G.M.T. This was one of a series of hours during which, by international agreement, the magnetographs
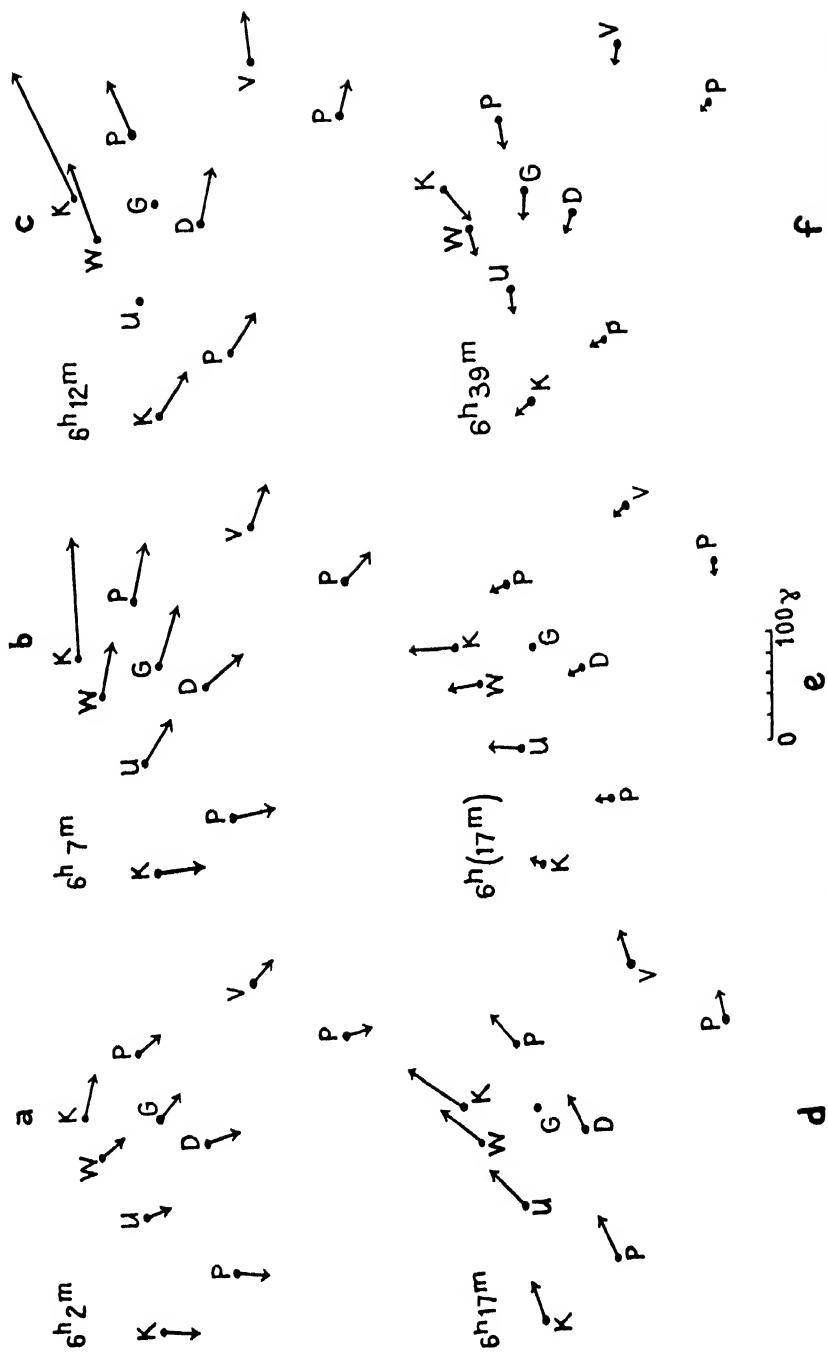
FIG. 27. Vectors of the horizontal component of the magnetic disturbing force at five epochs (G.M.T.) during a disturbance on February 28, 1896, at ten European stations. (After A. Schmidt)

at several observatories were run specially quickly, so that the movements were shown on an enlarged time-scale, enabling them to be more accurately timed. The observatories concerned in the figures were Kiel, Wilhelmshaven, Potsdam, Utrecht, Göttingen, Kew, Darmstadt, Paris, Vienna, and Pola. For D and H the mean for the hour was subtracted from the means for certain intervals of 1 minute (obtained as the average of 12 readings at 5-second intervals), and the corresponding horizontal disturbing vectors are shown in the figures.

Fig. 27 (a–c), for 18^h 2^m, 7^m, and 12^m, shows clearly the approach and recession of what seems to be a gradually strengthening current-vortex, which in Fig. 27 (d) (for 17^m) is still strong but more distant. Fig. 27 (e) represents the same force-distribution as Fig. 27 (d), after the subtraction of a constant force, the same for all stations; this force is supposed to represent approximately the influence of the distant vortex; it shows that the irregularities apparent in Fig. 27 (d) can be ascribed to a new small vortex, to the north; this vortex is distinct from the other, because its vectors converge, while those of the former diverge. Fig. 27 (f) (for 39^m) shows another case of a vortex which tends to make the north poles of the needles converge.

In considering these vector systems Schmidt supposed the direction of the current to be approximately perpendicular to the force vectors, the error being smaller, the more regular and intense the vortex. In Figs. 27 (b, c, f), which indicate a well-developed adjacent vortex, the current-lines will be almost circular.

9.18. Rotational changes of the magnetic vector during storms. Another interesting feature of magnetic disturbance was discovered by Sangster [25]. He examined the magnetic curves for D, H, and Z at Greenwich during the principal magnetic disturbances occurring in the years 1900–7, for which the traces were reproduced in the annual volumes of Greenwich observations. Selecting some of these disturbances for detailed study, he constructed diagrams of the changes in F, the total magnetic force; these diagrams show the change in *direction* of the disturbing vector ΔF, as seen by an observer looking along the mean direction F; thus they indicate the changes as projected on a plane perpendicular to F; the two components of these changes are $H\,\Delta D$ and $F\,\Delta I$ (ΔD and ΔI being reckoned in circular measure); $F\Delta I = (\cos I)\Delta Z - (\sin I)\Delta H$. At various points along the curve the values of ΔF, the magnitude of ΔF, are also indicated.

Fig. 28 A (1 and 2) refers to the disturbance on 1903 October 12, from 18^h onwards; points on the curves were measured at 5-minute intervals
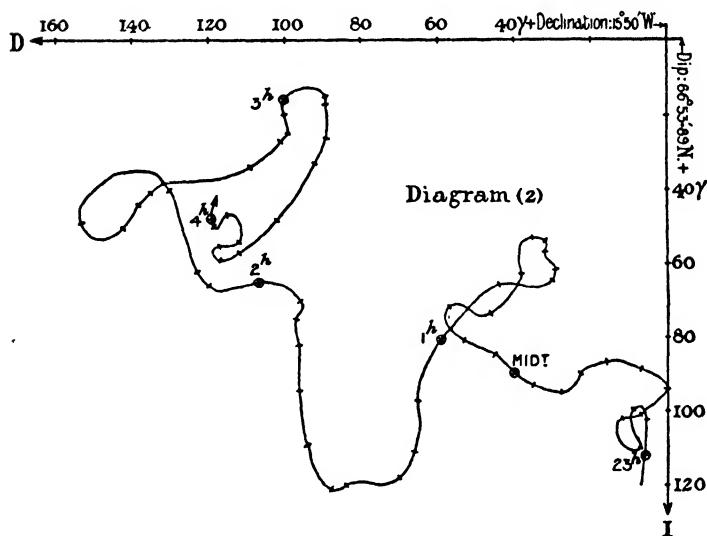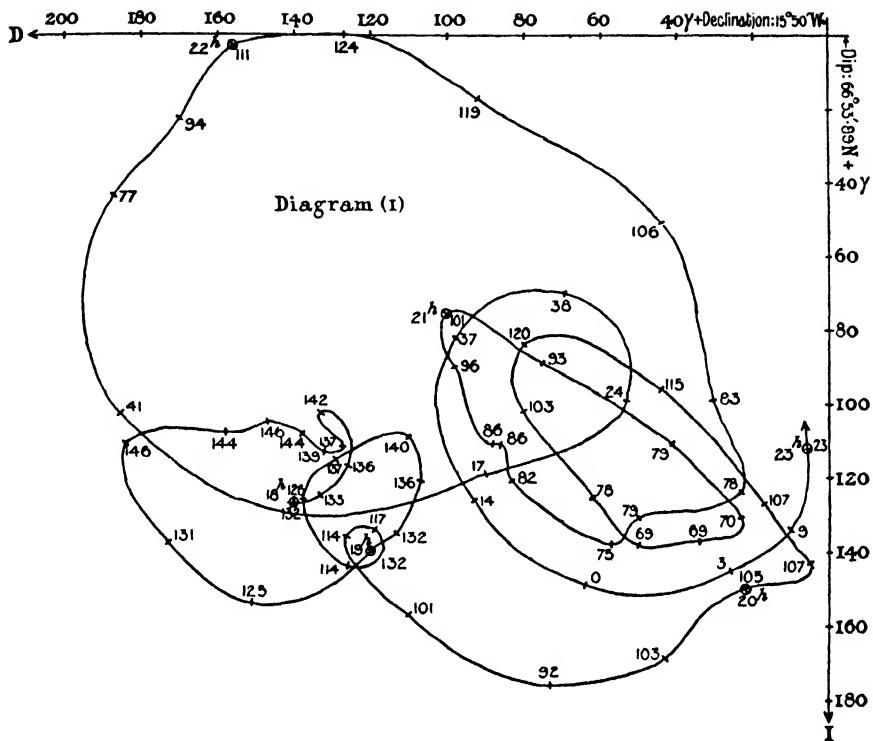
FIG. 28 A. Variation in direction of the total force as viewed towards the north-seeking end of a freely suspended dipping needle, Greenwich, 1903 October 12, hours 18–23 (diagram 1) and following 5 hours (diagram 2). (After R. B. Sangster)

(the small changes due to temperature variations of the recording instruments were ignored). Fig. 28 A (1), for the period 18^h to 23^h, shows that the movement of the end of the *F* vector was almost wholly rotatory and anti-clockwise. The successive convolutions vary greatly in their size and rate of description, but the long continuance of the anti-clockwise direction of the rotation is evident. Fig. 28 A (2) shows the curve from 23^h onwards; here the movements are less regular, and
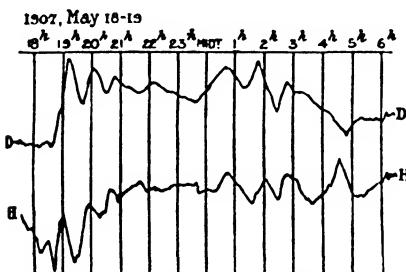
Fig. 28 B. Greenwich magnetograms, *D* and *H*, 1907, May 18–19

sometimes their direction is definitely clockwise. An examination of the ΔF values shown on the curve reveals that the end of the *F* vector moves nearly in a horizontal plane: thus the movements shown in Fig. 28 A (1 and 2) are reduced projections of this horizontal motion, on the plane normal to *F*.

The *direction* of motion of the end of the *F* vector can, however, usually be seen from the *D* and *H* traces without drawing such curves as Fig. 28 A (1 and 2). Fig. 28 B shows a good example of the two directions of rotation; it refers to the disturbance of 1907 May, 18^d 19^h to 19^d 5^h (Plate IV of the 1907 Greenwich volume). In this figure *D* and *H* increase downwards, and an increase of *D* denotes a movement of the end of the *F* vector westwards or to the left, while an increase of *H* means a northward movement. When the end of the *F* vector is moving clockwise, *D* is decreasing while *H* is above its mean, and increasing while *H* is below the mean; and vice versa when the motion of the end-point of *F* is anti-clockwise.

By a statistical study of the occurrences of clockwise and anti-clockwise rotation of the end of the *F* vector, Sangster found that the latter predominates before midnight and the former after midnight. Leaving out of account some of the printed Greenwich traces (1900–7) for which the changes were too rapid to permit of exact time comparisons between *D* and *H*, he examined seventy-two disturbances, and noted whether in each half-hour, beginning or ending at an exact Greenwich hour, the

rotation was clockwise ($+$) or anti-clockwise ($-$); only those cases were counted in which at least half a circuit was described by the end of the F vector; in a few cases $+$ and $-$ rotations occurred in the same half-hour, and each was reckoned to the credit of that half-hour. Adding up all the occurrences of rotation of either sign in each half-hour from all the disturbances, a frequency table giving the daily distribution of such rotations was obtained, as follows: blank denotes absence of cases of rotation.

	Winter		Equinox		Summer		Year	
	$+$	$-$	$+$	$-$	$+$	$-$	$+$	$-$
Midnight to 2^h	6	13	19	6	13	4	38	23
2 to 4^h	4	4	19	2	15	2	38	8
4 to 6^h	6	1	12	3	9	1	27	5
6 to 8^h	..	..	3	1	4	..	7	1
8 to 10^h	2	3	3	2	2	..	7	5
10 to noon	..	2	..	2	..	1	..	5
noon to 14^h	..	..	..	5	..	3	..	8
14 to 16^h	..	3	..	6	4	2	4	11
16 to 18^h	..	18	..	20	..	8	..	46
18 to 20^h	..	30	..	35	.	9	..	74
20 to 22^h	4	30	..	37	..	18	4	85
22 to midnight	6	28	8	32	1	13	15	73

In considering the table it must be remembered that disturbance is most frequent at Greenwich between 22^h and 24^h, so that the Greenwich records give more night than day disturbances. The table shows that from 10^h to 24^h almost all the rotations were anti-clockwise, while from midnight to 6^h clockwise rotations predominated; from 7^h to 9^h *only* clockwise rotations occurred.

Another example of rotatory movements of the magnetic field vector during a magnetic storm has been given in a stereogram [2.79]. ·[a]

9.19. Birkeland's first memoir. At about the same time as Schmidt, Birkeland began his very extensive and important work on magnetic storms. Realizing the need of more polar magnetic data, he organized Arctic expeditions, first in 1899–1900 to Hallde· near Bossekop, in the far north of Norway, and again in 1902–3, when four magnetic observatories were maintained for several months, at Kaafjord (Bossekop), Dyrafjord (Iceland), Axeloen (Spitzbergen), and Matotchkin Schar (Nova Zembla). He discussed the data thus obtained, along with data from many other observatories, both for these periods and for the polar year 1882–3, in three important memoirs, of 1901, 1908, and 1913.

In his first memoir [G 97] he discussed the data obtained during the international polar year 1882–3, as well as his own Norwegian data of 1899–1900; he found that frequently the magnetic variations were such as would be produced by concentrated currents .flowing along the auroral zone, at a height which he estimated as between 100 and 300 km., the mean height being near 100 km. The current-intensity, at times of strong disturbance, he estimated as 1,000,000 amperes. The evidence for these conclusions was that at stations north and south of the auroral zone the vertical force is disturbed in opposite senses, whereas the horizontal force disturbances are in the same sense, and such as to indicate that the disturbing cause is above the earth; thus, in Fig. 29, let A represent the cross-section of a horizontal electric current flowing (*into* the paper) above and parallel to the ground (CD). At C
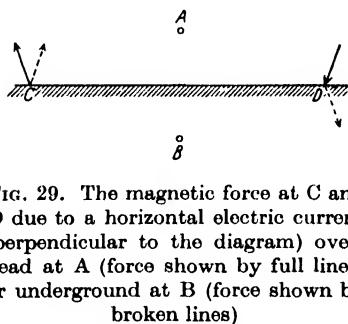
Fig. 29. The magnetic force at C and D due to a horizontal electric current (perpendicular to the diagram) overhead at A (force shown by full lines) or underground at B (force shown by broken lines)

See also p. 749.

and D the magnetic force due to A is shown by full arrows; dotted arrows show the corresponding magnetic forces due to a parallel current flowing under the ground, at B, into the paper. The currents A and B affect the vertical force in the same sense, but their effects on the horizontal force are opposite; if the current at B were reversed, its horizontal magnetic force at C and D would be the same as that of the overhead current at A, but the vertical magnetic force would be opposite; thus a field due to an underground current can be distinguished from one due to an overhead current, if it is assumed that the current flows along a line. The magnetic changes observed by Birkeland near the auroral zone were of the type due to A, or the opposite (corresponding to a reversal of this current). The currents were found to be sometimes eastward and sometimes westward. The intensity and height of the current at A could be roughly calculated from the magnetic disturbance observed at suitable points C and D on opposite sides of the current-line.

For example, he considered (loc. cit., p. 27) several cases in which at Jan Mayen in Spitzbergen, and Bossekop in Norway, the horizontal force was equally perturbed and the disturbances in the vertical force were equal but in opposite senses. He supposed these effects to be due to an electric current flowing midway between the two stations, at a height provisionally assumed to be 200 km., and in a direction perpendicular to the horizontal force perturbation. In a particular case

(Fig. 30), on 1883 January 2, at $1^h 40^m$ G.M.T., this direction was such that the horizontal distance of the current from each station was 280 km.; the two stations are 1,220 km. apart. From the vertical-force
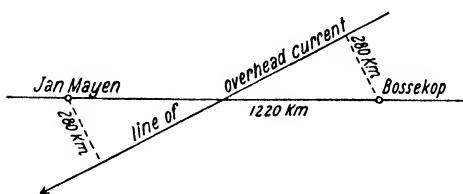
FIG. 30. The situation, in plan, of an overhead horizontal electric current, inferred from the vectors of the disturbing magnetic force at Jan Mayen and Bossekop

perturbation (150γ) the current was estimated to be 317,000 amperes. Other cases may be quoted:

1882	Oct. 15.	1^h	.	.	.	556,000	amperes
1883	Feb. 1.	21^h	.	.	.	350,000	,,
1883	July 15.	23^h	.	.	.	520,000	,,

As a check on the assumed height of the current, Birkeland considered two perturbations in which the current appeared to pass nearly above Bossekop—as shown by a negligible vertical-force perturbation and an exceptionally great change in H—while at other stations its effects were of the same order as those in the preceding cases. His results were:

1882	Nov. 15.	1^h	.	.	.	497,000	amperes
1883	July 1.	1^h	.	.	.	983,000	,,

These estimates are too low if, as is likely, the currents are actually spread out over a layer extending many kilometres on either side of Bossekop. Since the estimates, however, are greater than the former ones, he inferred that the height of the currents is somewhat lower than 200 km. (see, however, 22.21 and Fig. 22.5, p. 749).

Throughout his studies of magnetic disturbances Birkeland systematically considered what type of horizontal electric currents flowing above the earth would give rise to the observed magnetic perturbations. This plan, which has also been followed (by mathematical or approximate methods) in regard to the S_q, L, and averaged D fields in Figs. 7.15, 7.16, 8.4–8, and 9.21–6, is of great service in affording a simple representation of the magnetic field at numerous stations, whether or not the field is produced by such currents. The currents being assumed to flow above the earth, their direction can be inferred from the D and H perturbations alone; the Z perturbations, when

known, always confirmed the assumption that the main currents were above and not below the ground.

In his first memoir Birkeland considered all the greater perturbations registered at Bossekop and Hallde in 1899–1900, for which the disturbances in the various elements could be fairly definitely measured by reference to their values during a neighbouring period of magnetic calm. In his calculations the perturbing force in the horizontal plane ranged up to 100γ or a little more; at his two magnetic observatories the larger disturbances were not properly registered on the sheets. But the perturbations recorded during the polar year 1882–3 ranged up to 700γ or even more. Using the records then obtained for Bossekop, Fort Rae, Jan Mayen, Pavlovsk, Christiania, Copenhagen, Potsdam, Paris, Greenwich, and Toronto, Birkeland endeavoured to form a connected idea of the current-systems involved. He concluded that both the lesser and greater perturbations observed in the polar regions are due to overhead electric currents, which flow partly along the auroral zone (where they are very concentrated), and spread outwards from the zone so as to traverse a great part of the globe; they do not rotate with the earth, being fixed relative to the sun, apart from occasional changes of direction, which often occur rather suddenly. He drew a diagram of these currents as they exist at Greenwich midnight; this is here shown as Fig. 31; it depicts the northern hemisphere viewed from above the north pole, and is placed so that the midday meridian is directed downwards. It will be seen that the system bears some, but only a slight, resemblance to the current-diagram shown in Fig. 26, which is derived from the *average* daily magnetic variations in polar regions. The current-lines are not closed to form complete circuits, and in some places they cross one another. To a certain extent, however, they support the conclusion stated in §§ 4, 5, that the *type* of the magnetic disturbance field is fairly constant over a wide range of intensity; the disturbances on which Birkeland's diagram is based are many times as intense as those from which Fig. 26 was drawn. A further difference is that Fig. 26 refers to polar magnetic changes averaged over many days, while Birkeland's perturbations often lasted only for a few hours. He found that such perturbations usually corresponded to 'pulsations' of the current-systems of his figure; these pulsations appear to affect both the intensity of the currents and, to some extent, their situation. From the stations named, Birkeland found that at about 21^{h} G.M.T. (or about 22^{h} at Bossekop) the perturbations occurred notably earlier at the more easterly than at the westerly stations; he

suggested that the perturbing system was travelling westwards ('oscillates from east to west', p. 34), at times when the current-system was suddenly being intensified; the velocity was estimated as about 100 km./minute, though it seemed not to be uniform at different times
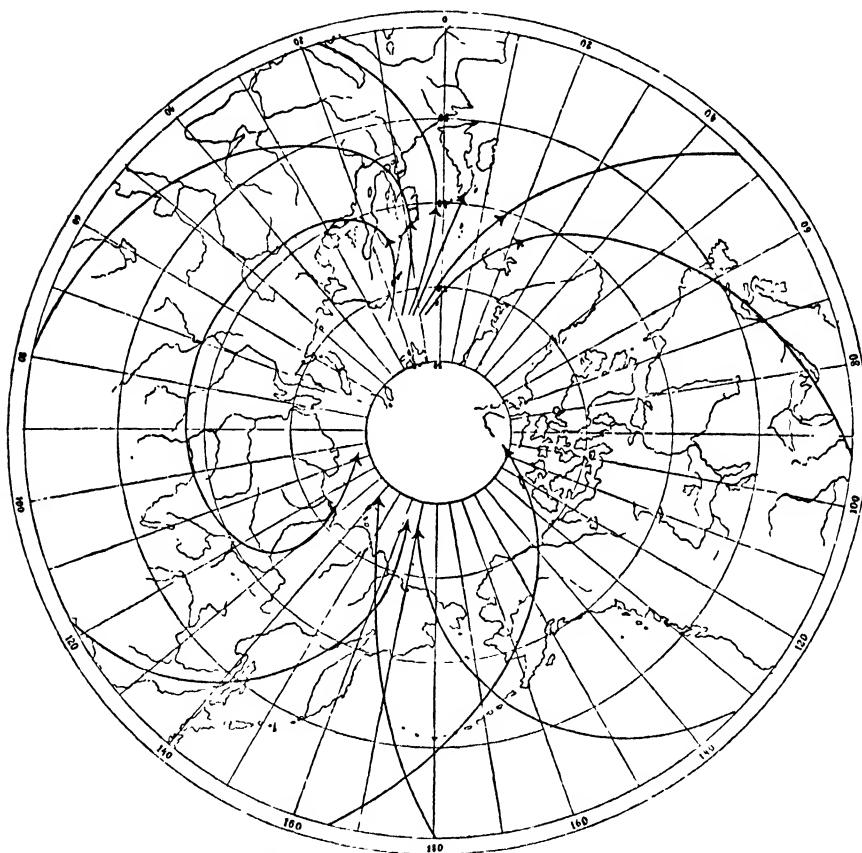
FIG. 31. Birkeland's diagram of the overhead electric current-system at the epoch of Greenwich midnight, during a typical polar magnetic storm

of the day—especially at 8–10^h G.M.T. it was abnormal. But in his discussion of these points his statements are sometimes rather obscure.

9.20. Birkeland's classification of magnetic storms. In his second memoir (1908) Birkeland classified magnetic perturbations into five groups, each having comparatively well-defined characteristics [G 97 a]. He regarded them, however, as having 'a certain genetic connexion' with one another, and therefore as being not altogether separate phenomena

3595.30   Y

'It appears that magnetic storms of any considerable strength are most frequently of a kind in which the force increases towards the poles. It also appears, however, that it is not unusual to find perturbations that are best developed and most powerful at the equator . . . these perturbations in the region of the equator act principally upon the horizontal intensity. . . . As regards the lower latitudes, the circumstances of the perturbation often exhibit symmetry both with respect to the magnetic axis and to the equator. Such perturbations we have chosen to call equatorial perturbations. Of these there are again two kinds possible, namely, such as produce an increase in the horizontal force, and such as produce a diminution' (p. 62).

The first he called *positive equatorial perturbations*, and the second, *negative equatorial perturbations*. 'The two have quite a different character and course. The positive equatorial perturbation in particular is strongly characterized.'

His next two classes were the *polar positive* and *negative* perturbations, in which the most powerful forces are found in the polar regions, and the forces decrease very rapidly in strength with descent to lower latitudes. The signs positive and negative refer to the cases in which the horizontal intensity is respectively increased or decreased, near the auroral zone, corresponding to an easterly or westerly overhead current. The negative and positive storms are very similar except in sign, but the negative ones are usually the stronger and show the more disturbed magnetic curves, especially in regard to rapid serrations; this suggests that the disturbing cause (taken to be external current-systems) must approach the earth comparatively closely. The disturbance field is such as could be produced by a large overhead horizontal current-system consisting of a concentrated current along the auroral zone, completing its path by dispersed currents to the north and south.

Fig. 32 shows a diagram given by Birkeland to illustrate the 'typical field' of a polar elementary storm. The direction of the central arrow through C is to be interpreted as easterly in the case of a positive polar storm, and westerly for a negative polar storm. The dotted lines are the current-lines, the full lines are lines of the horizontal disturbing magnetic vector. The lower part of the diagram shows the latitude-distribution of the perturbation in the upward vertical force, along the meridian through the centre C of the whole current-system. The upward vertical force is increased on the right of the central arrow, which corresponds to the north or south of the zone, for positive and negative polar storms respectively. The current-flow near the central arrow corresponds to flow along the auroral zones to east or to west. Birkeland discussed this type of polar current-system as follows.
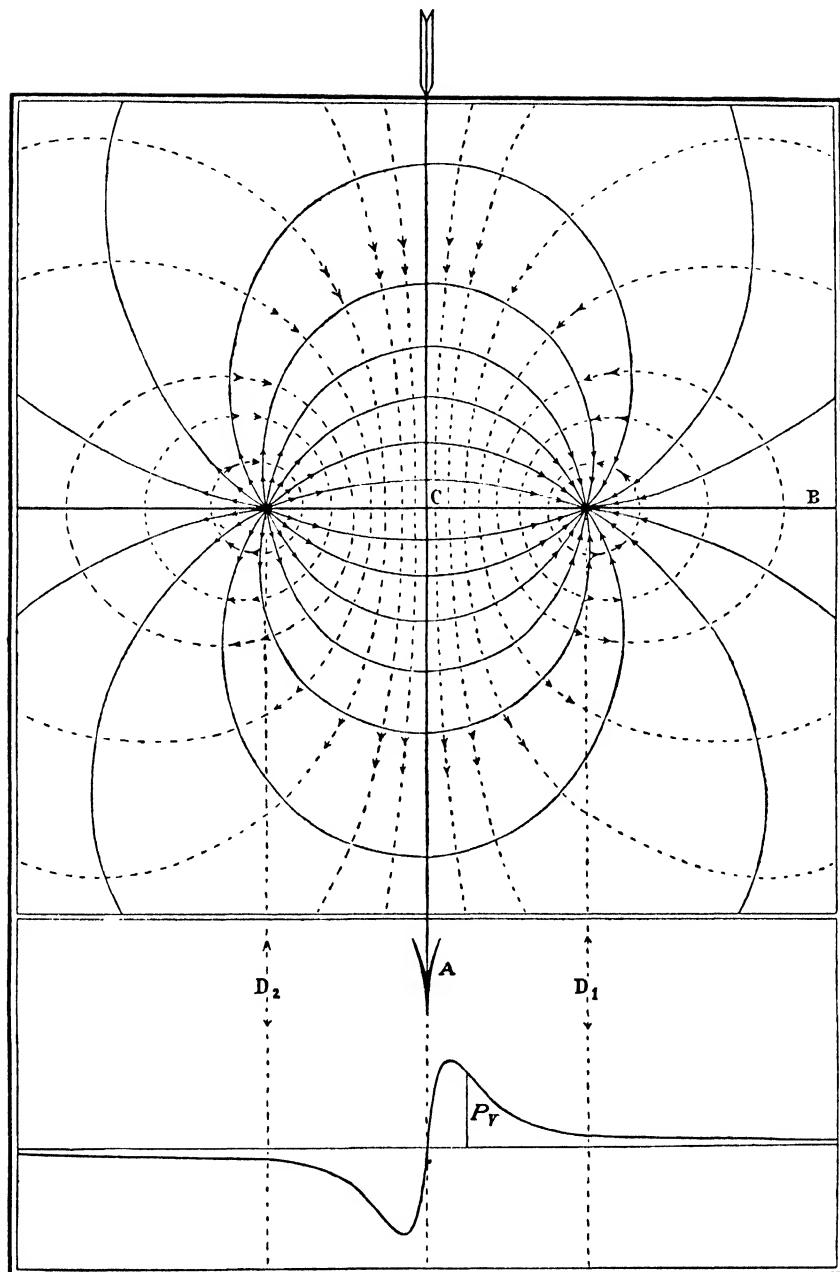
FIG. 32. Birkeland's diagram of the overhead electric current-system (broken lines) that might produce the disturbance-field of his 'polar elementary storm'; the corresponding lines of horizontal magnetic force for this disturbance-field are shown by the full lines. The lower part of the diagram shows the variation, along the line BC, of the vertical component of the disturbing force

'These two principal systems, the negative and the positive polar perturbation systems, rarely occur quite alone. As a rule they occur simultaneously, but in different districts. It appears that they always, on the whole, are grouped in the same manner in relation to the sun, and in the following manner. On the morning and night sides of the globe there is always a powerful, negative polar system of precipitation, generally fairly extensive, in which the principal axis of the system falls, as a rule, along the auroral zone. The negative system continues westwards on to the afternoon side, but here the principal axis of the system turns northwards to the districts north of the auroral zone, and it looks as if the system also as a rule would be continued westwards until it joined the negative system on the morning side. . . . The positive polar system develops along the auroral zone, most strongly in the southern part of the zone. It may sometimes be of very considerable extent, but as a rule is much smaller than the negative system. In this way there will be a boundary station in the auroral zone, as a rule upon the evening side, which will be situated between the positive and negative systems. Thus, while at the stations on the afternoon side in the auroral zone, the positive storm is the principal phenomenon, and on the night side the negative, and the perturbations here occur with great distinctness and with well-defined deflexions in a positive or negative direction, as the case may be, at this boundary-station now one system, now the other, will prevail, causing the deflexions in horizontal intensity to be at one time positive, at another negative' (p. 445).

'In detail, however, we shall be able to find the perturbation conditions somewhat different from those we have here described as the typical. The forces will always . . . in the extended negative area of precipitation, concentrate themselves about one or several storm-centres. At the same time, the negative systems that occur at the other places will more or less disappear. Frequently there is a single, comparatively very limited, negative system of precipitation, while the rest of the negative current-circuit has practically disappeared. This has often proved to be the case at about Greenwich midnight.' [It is to be remembered that Birkeland in the main was discussing the perturbations that were selected from the registers of his four Arctic observatories, which would be most disturbed at particular local times, and therefore over a certain range of Greenwich time.] 'At about this hour, we very frequently find a powerful, well-defined, and comparatively very limited negative system in the north of Europe, while at other places round the Arctic zone, no negative systems of precipitation are apparent, as far as we can see from our observation-material. For this reason the storms that occur at this time exhibit particularly simple areas of perturbation' (p. 446).

The current-systems were supposed to move westwards, relative to the earth, this motion being interpreted as due to the earth's rotation, the current-systems remaining comparatively fixed relative to the sun. But deviations from this rule were often noted, the current-system

sometimes even moving eastward. This was taken to be due to changes in the region of precipitation of solar corpuscles into the atmosphere, due to the varying inclination of the earth's magnetic axis relative to the sun.

Birkeland stated that 'It is especially the positive and negative polar storms, and the positive equatorial storms, that are most frequently met with' (p. 439).

'The chief peculiarities of the positive equatorial storm are as follows:

'Everywhere in low and medium latitudes, positive perturbing forces are met with in the horizontal intensity, while at the same time in declination no deflexions, or only very small ones, are found. In the vertical intensity, only small perturbing forces appear . . . we find the strongest perturbing forces in the equatorial regions, while the perturbing forces decrease in strength with increasing distance from the magnetic equator' (p. 439).

'The deflexions in horizontal intensity always increase at the beginning of the storm rather rapidly and to a certain height, after which the perturbing forces remain more or less constant in strength for a long period.' (His 'two best' examples of positive equatorial storms suggest that by 'long period' Birkeland here meant about six hours.)

'In the horizontal-intensity curve, there are always a number of very characteristic serrations, which are found again at all the stations situated in low and medium latitudes, and these serrations appear at any rate very nearly simultaneously all over the globe. . . . If, on the other hand, we approach the auroral zone, the perturbation-conditions alter to some extent. We also find in declination deflexions like those in horizontal intensity. A peculiar impulse at the beginning of the perturbation, which was less noticeable in lower latitudes, now comes out distinctly, this being that the deflexions in horizontal intensity are not first in a positive direction, but in a negative; and the perturbing force oscillates here, at first quite distinctly through a more or less considerable angle. This condition is most distinct in the immediate vicinity of the auroral zone. Here, too, we find again serrations to some extent similar to those at southern stations, but often considerably larger.'

'Very frequently, perhaps as a rule, the positive equatorial storm is interrupted by the breaking in upon it of a polar storm' (p. 440). 'It is this feature that we continually find repeated, namely, that when the equatorial storm has lasted for some hours, polar systems appear' (p. 70); 'this equatorial perturbation often comes as a precursor of polar storms; and indeed, we have really never met with an entire perturbation of this kind with which there have not, within the same period, been polar storms' (p. 79).

'We thought of showing two more types of perturbations, namely, the negative equatorial storm and the cyclo-median storm' (his fifth type). 'We have, however, only a few examples of these among our observations.'

'The negative equatorial storms are most powerful in the region of the equator, where the perturbing forces in horizontal intensity are negative.

The forces that occur in the negative equatorial storms are also considerably greater than those found in the positive. Among our observations we have found only examples of negative equatorial storms, which occur simultaneously with polar storms, and it is perhaps doubtful whether this type of perturbation on the whole can occur alone. We have not sufficient material, however, for the formation of any well-founded opinion on the matter' (p. 447).

Birkeland's fifth type of elementary perturbation was called by him a *cyclo-median perturbation* or storm, because it was such as could be ascribed to a closed (or cyclic) current-circuit of limited size not far above the earth, and situated in middle (or *median*) latitudes; it was 'not a polar disturbance at all'. The direction of flow of the supposed current, viewed from above, was counter-clockwise.

'We have only a few instances of such perturbation-fields that can be characterized as rather well defined. We believe the perturbation of the 6th October is a storm of which the field of force should be explained as the effect of such a cyclo-median system. In our discussion of the compound perturbations, we have also several times come across fields that would naturally be due to cyclo-median systems, but in which nothing could be decided, owing to the complicated character of the storm . . .' (p. 448).

He then gives references to further cases of the kind.

Birkeland's method of computing current-intensities from the magnetic displacements, hour by hour, has been applied also by Goldie [22, 23] to certain groups of storms; he thus devised a composite electric current-system for the northern part of the northern hemisphere. According to his diagram a large part of the current entering the auroral zone (where the current-concentration is great) from lower latitudes does not appear to return southward, and 'must therefore be presumed to be dissipated in the auroral zone and mainly in the region from about 20 h. to 2 h. The picture thus derived from the magnetic data is one which requires a kind of electrodeless discharge in the auroral zone in order that the various parts of the system may be fitted together. From the mathematical point of view, however, the entry of negative electricity into the auroral zones from outside the earth's atmosphere could satisfy the rest of the picture.' This question of the actual situation of the current-systems responsible for the D field is discussed in Chapter XXV.

9.21. Discussion of Birkeland's classification of magnetic storms. In a review of Birkeland's work, Chree pointed out that the disturbances considered by Birkeland included no great magnetic storm, and that, indeed, most of his perturbations would not be regarded as

magnetic storms at all, in the ordinary usage of the term [G 15]. This was due to the fact that his Arctic expedition coincided with a period which magnetically was fairly quiet.

Although Birkeland regarded his various types of storms as having a certain connexion with one another (owing, as he believed, to their supposed origination by solar electrons), he seems not to have realized that a great magnetic storm is a unitary phenomenon, going through regular phases. This seems evident from our discussion of storms in §§ 1–5, in which the changes are separated according to storm-time and local time. It seems that Birkeland's positive equatorial storm corresponds to the first phase of the non-polar part of a great magnetic storm, and his negative equatorial storm to the second, major, phase; he recognized that negative equatorial storms can be more intense than positive ones. He found only a few examples of them among his data, and since the changes during the major phase of a storm are slower and more gradual, the weaker the storm, it is natural that the negative perturbations which occurred in his data (drawn from a not very disturbed period) would not be conspicuous.

His positive and negative polar storms agree with the average character of disturbance found by Chapman in the polar regions, and seem to be clearly part of a single phenomenon, waxing and waning in unison with the non-polar disturbance field. This is specially notable at times of great storms, when aurorae and the polar magnetic disturbance-field become extended to unusually low latitudes.

Birkeland's typical disturbances often extend over only a few hours, and rarely for as long as a day, whereas magnetic storms are often active for a day, and sometimes even longer. This tends to confirm the view that he split up the perturbations artificially into small portions that ought really to be considered together.

Hence his first four classes of storm are concluded to be separate aspects or phases of one phenomenon. All four must be explained together, as resulting from a single primary cause, though the different phases, and the effects in different regions, may be produced by this one cause in different ways. In his terminology, a magnetic storm begins with a positive equatorial perturbation, succeeded by a negative equatorial perturbation, during which, over the polar caps, there are positive and negative polar storms, the positive ones being centred on the evening side of the earth, and the negative ones on the morning side.

Birkeland's cyclo-median storm, to judge from the one definite example which he gives, is a fleeting disturbance of the curves occurring

on an otherwise quiet day. From his description it appears to be an S_q augmentation (10.3), such as has recently been found to be associated with a particular type of solar eruption and with radio fade-outs over the day hemisphere. Birkeland seems to have been the first to recognize the existence of this type of variation of the earth's field as a separate kind, quite distinct from the ordinary magnetic disturbance which has such marked polar relationships. The elucidation of the magnetic disturbances associated with solar flares and radio fade-outs is due to Fleming and McNish.

9.22. Remarkable magnetic storms. Since photographic recording of the magnetic field was begun, in 1857, a number of lists of magnetic storms have been published, which, although they may be based on the records of only one observatory, are practically valid for the whole earth as regards the times of commencement. Such lists were given by Moos (1872–1904, Bombay) and Ellis (1848–97, Greenwich). A valuable comparison of magnetic storms and sunspots, 1874–1927, is contained in the Greenwich Photoheliographic Results for 1927; the first catalogue gives 60 greater storms, for which the ranges were greater than 60' in D or 300γ in H or Z, and the second catalogue gives 343 smaller magnetic storms with ranges between 30' and 60' in D, and between 150 and 300γ in H or Z (see 11.16).

The most violent storms since 1857 are as follows: the ranges in D, H, and Z are indicated where possible.

1859 August 28–September 7 (see § 23).

1872 February 4: $\Delta H > 960\gamma$ at Bombay, and aurora seen there.

1882 November 17–21: $\Delta D = 115'$, $\Delta H > 1,090\gamma$, $\Delta Z > 1,060\gamma$ at Greenwich. This storm synchronized with the passage of the largest sunspot group of that cycle (greatest projected area 4,700 millionths of the sun's disk).

1903 October 31–November 1: $\Delta D = 3° 6'$, $\Delta H > 950\gamma$, $\Delta Z > 950\gamma$ at Potsdam.

1909 September 25: $\Delta D = 3° 30'$, $\Delta H > 1,500\gamma$, $\Delta Z > 1,100\gamma$ at Potsdam. This storm lasted only 10 hours.

1921 May 13–16: $\Delta D = 3° 19'$, $\Delta H = 1,060\gamma$, $\Delta Z = 1,100\gamma$ at Potsdam. Aurora was seen at Samoa, in $13·8°$ S. latitude (see Fig. 34).

1938 April 16: $\Delta D = 5° 28'$, $\Delta H = 1,900\gamma$, $\Delta Z = 600\gamma$ at Potsdam (Niemegk). Records of this storm are reproduced in Figs. 33 A and B. The violence of this exceptional storm is illustrated by a comparison of the two diagrams: Fig. 33 A shows the main variations on such a small scale that the storm seems practically over by 10^h

G.M.T. Actually 'the storm was then still continuing, as is shown by Fig. 33 B, which represents the variations from this time onwards on the usual scale. The records of the day preceding the storm are also shown, and the comparison of the levels of the records before and after the storm (right end of Fig. 33 B) gives a good example of the 'post-perturbation': X has decreased by 60γ, Y has
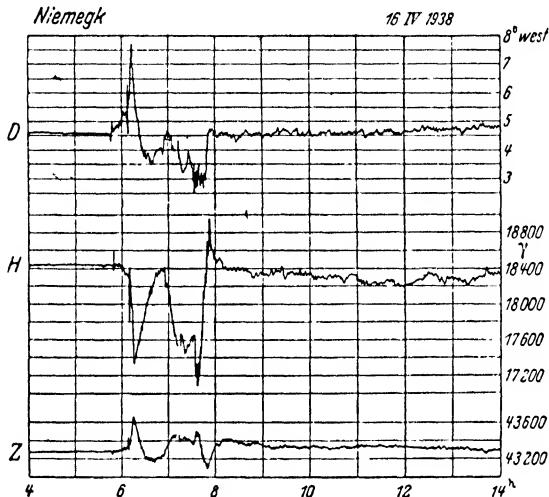
FIG. 33 A. The magnetic storm of 1938 April 16, as recorded at Potsdam (Niemegk), from 4^h to 14^h Greenwich Mean Time, by special insensitive variometers for western declination and horizontal and vertical intensity

increased by 20γ and Z by 35γ. The ratio $(20:60 = 1:3)$ of the after-effects in Y and X is practically equal to $\tan 19°$, and $19°$ west of north is, in fact, the azimuth, at Potsdam, of the plane through the magnetic axis of the earth. The small after-effect in Z compared with X ($35:60 = \tan 30°$) is due to the effect of currents induced in the earth, which (numerically) enlarges ΔX and diminishes ΔZ; if the vector of the post-perturbation were parallel to the magnetic axis, the ratio $\Delta Z : \Delta X = \tan 52° = 1.3$ would be expected.

Details of other storms are to be found in the reports of observatories and in notes published in various journals (such as 'Principal magnetic storms', quarterly in the journal *Terrestrial Magnetism*, and sometimes also in the *Meteorologische Zeitschrift*). A good collection of reproductions of Potsdam records has been given by Rössiger [27]. As to the biggest storms, most of the observatories have lost at least part of their records during these periods of exceptional interest. It is to be

hoped that the next great magnetic storm will at last find more observatories prepared with at least one set of insensitive wide-range instruments; although these are the most easily managed, and the most valuable during very disturbed periods, they are as yet rarely installed in observatories (pp. 54, 55).

Bauer [28] has drawn attention to a moderate world-wide magnetic
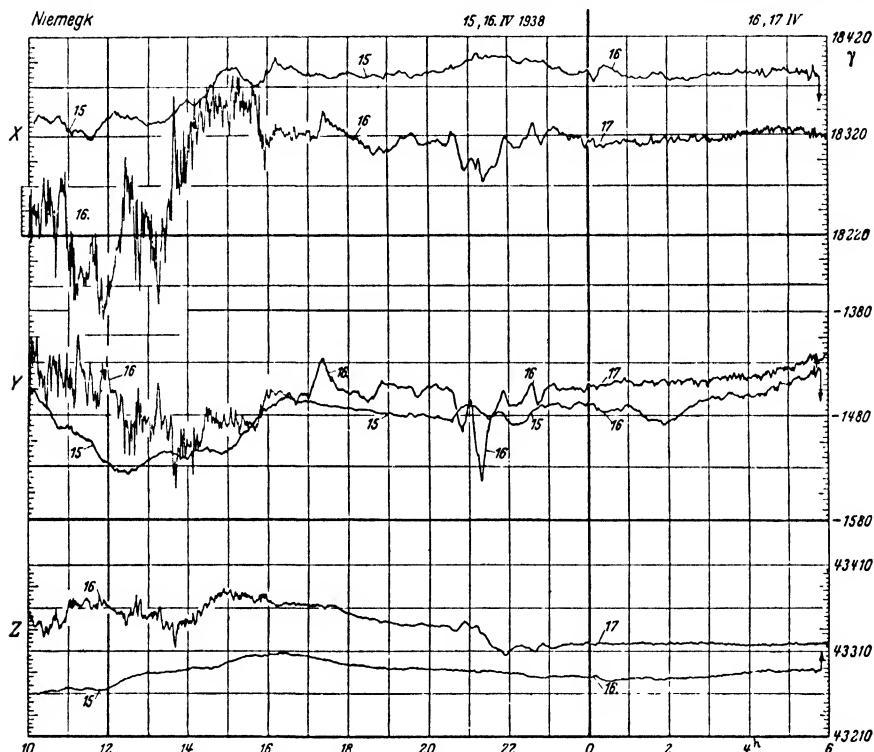
Fig. 33 B. The variations *after* and *before* the main phase of the same storm, as recorded from April 16, 10ʰ to April 17, 6ʰ G.M.T. and from April 15, 10ʰ to April 16, 6ʰ by the ordinary variometers for the north, east, and vertical components, showing, by the change of the general level, the effect of the post-perturbation. The violence of the main phase of the storm may be judged by comparing the hours 10 and 14, April 16, shown on Figs. 33 A and B; the variations in *H* are similar to those in *X*, and *D* corresponds to *Y*

storm which began, within a few minutes, simultaneously with the volcanic eruption of Mont Pelée, Martinique, which destroyed St. Pierre on May 8, 1902. There occurred, however, similar disturbances in the weeks before, so that the coincidence may be regarded as fortuitous, especially since no other coincidence of this kind has been observed. Earthquakes cause mechanical oscillations of the magnetic needles, but no certain magnetic effect has so far been found.
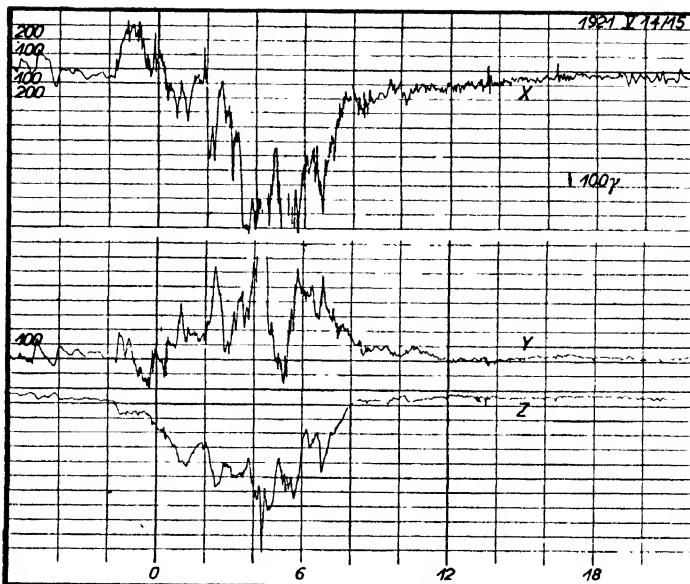
FIG. 34 A. The Potsdam records of the great geomagnetic storm of 1921 May 14
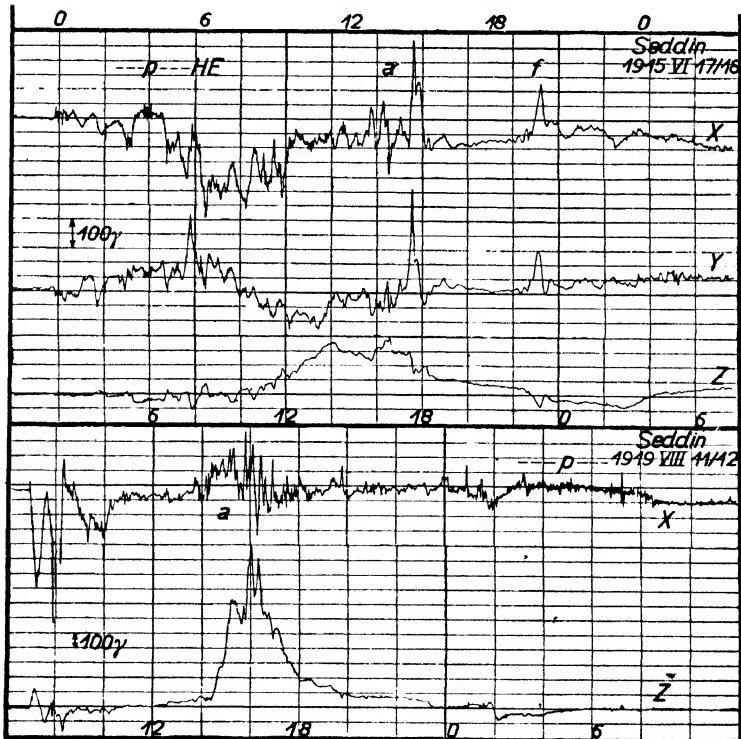
FIG. 34 B. The Seddin records of two great geomagnetic storms

9.23. Some historical magnetic observations. (*a*) In 1836 Gauss and W. Weber started the *Magnetische Verein*, a group of several European observatories taking eye-readings of declination and (later) horizontal intensities every 5 minutes for prearranged intervals of 24 hours, four per year. The results for six years, 1836–41, were published. It must be considered as a fortunate coincidence that 1837
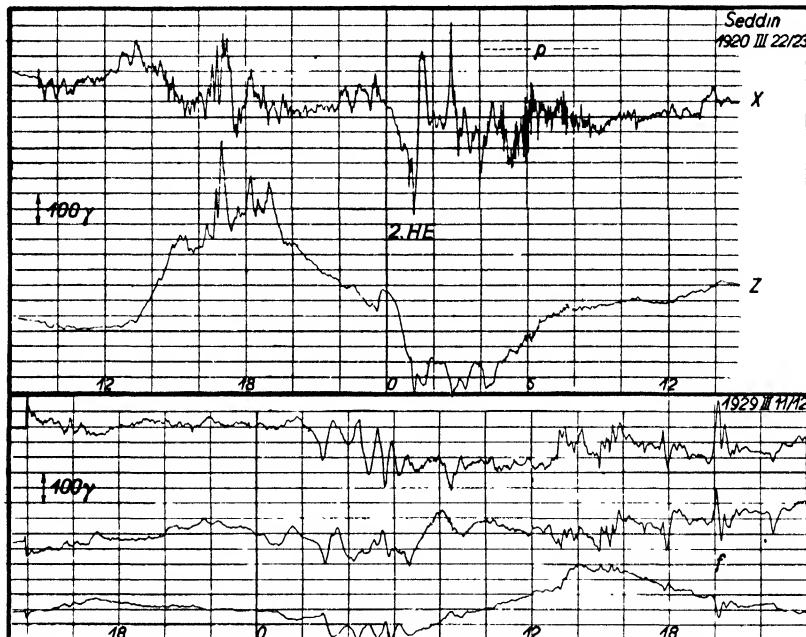
FIG. 34 c. The Seddin records of two great geomagnetic storms

was a year of maximum sunspot activity (annual mean sunspot-number 138), with correlated high magnetic activity, so that quite a number of these selected days happened to show considerable magnetic fluctuations which, in turn, increased interest in these phenomena. The interval between 10 p.m. Göttingen mean time, August 30–31, 1839, shows variations in Göttingen with a maximum range of about 35′ in *D*, and 160γ in *H*; it was accompanied by aurora observed in Upsala.

(*b*) Photographic recording (26.16) at Kew Observatory was begun in 1857 by Balfour Stewart. We quote [28 *a*] from his report 'On the great magnetic disturbance which extended from August 28 to September 7, 1859' (see also 13.1).

'During the latter part of August, and the beginning of September, 1859, auroral displays of almost unprecedented magnificence were observed **very** widely throughout our globe, accompanied (as is invariably the case) with

excessive disturbances of the magnetic needle. The interest attached to these appearances is, if possible, enhanced by the fact that at the time of their occurrence a very large spot might have been observed on the disk of our luminary. . . . In not a few instances telegraphic communication was interrupted, owing to the current produced in the wires; and in some cases this proved so powerful that it was used instead of the ordinary current, the batteries being cut off and the wires simply connected with the Earth. . . .
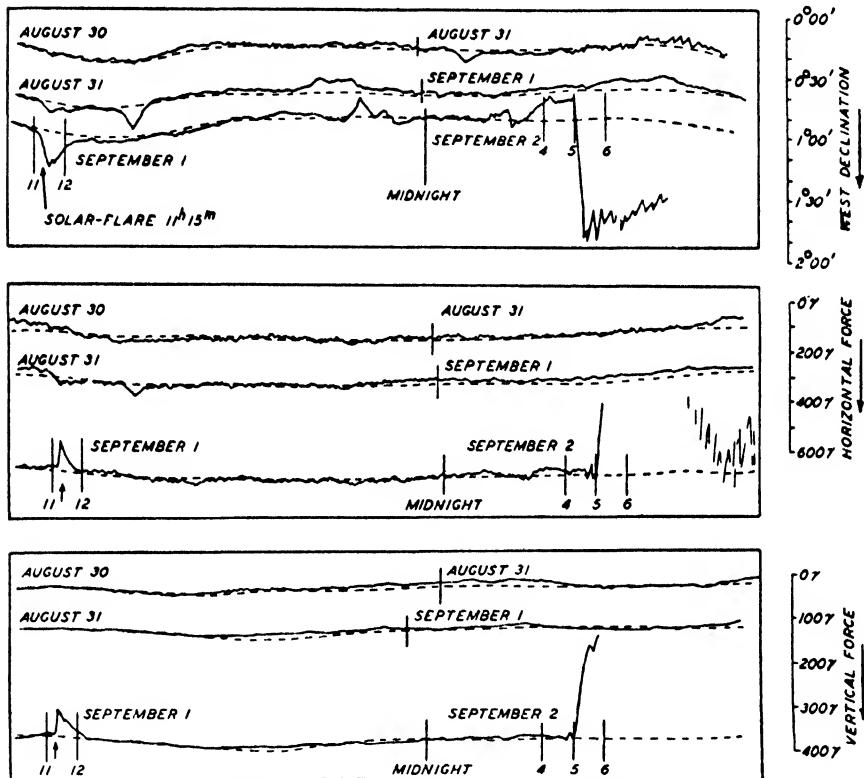
FIG. 35. Magnetograms, Kew, August 30 to September 2, 1859

We have two distinct well-marked disturbances, each commencing abruptly and ending gradually, . . . on the evening of August 28, and on the early morning of September 2. They correspond in time to the two great auroral displays. . . . It is impossible to state with accuracy what were the greatest departures from the mean value, as the curves for all the elements went beyond the sensitive paper; very approximately, they are estimated as 2° 20′ in declination, −0·04 of the whole in H [about 700 γ], 0·01 of the whole in Z [about 400 γ].' (Fig. 35.)

On 1859 September 1, at about $11^{\mathrm{h}}\,18^{\mathrm{m}}$ G.M.T., the well-known solar observer Carrington happened to be observing, by means of a telescope,

a large spot on the sun. The following extract is from his report (Fig. 36):

'The image of the Sun's disk was, as usual with me, projected on to a plate of coated glass . . . a picture of about 11 inches. I had secured diagrams of all the groups . . . when within the area of the great north group . . . , two patches of intensely bright and white light broke out. . . . My first impression was that by some chance a ray of light had penetrated a hole in the (shading)
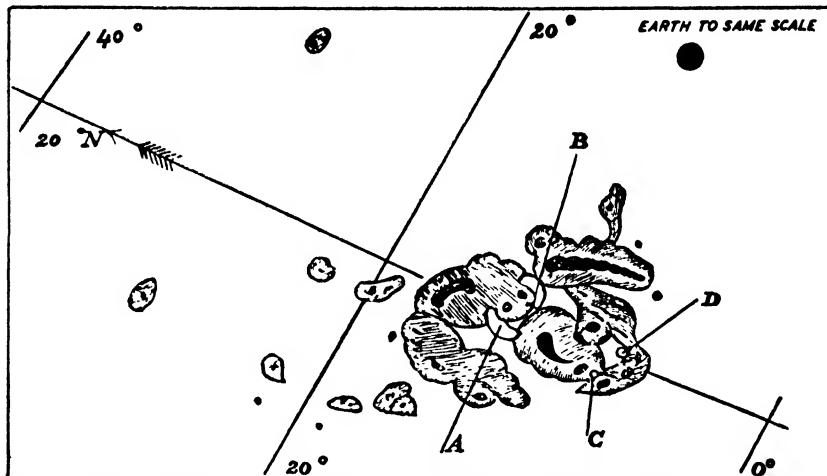
FIG. 36. Solar sketch, September 1, 1859, by R. C. Carrington

screen attached to the object-glass, for the brilliancy was fully equal to that of direct sunlight; but . . . by causing the image to move I saw I was an unprepared witness of a very different affair. I thereupon noted down the time by the chronometer, and, seeing the outburst to be very rapidly on the increase, and being somewhat flurried by the surprise, I hastily ran to call some one to witness the exhibition with me, and on returning within 60 seconds, was mortified to find that it was already much changed and enfeebled. Very shortly afterwards, at 11^h 23^m, the last trace was gone. . . . I was certainly surprised, on referring to the sketch finished before the occurrence, at finding myself unable to recognize any change whatever as having taken place. The impression left upon me is, that the phenomenon took place at an elevation considerably above the general surface of the sun, and accordingly, altogether above and over the great group in which it was seen projected. . . .' (Fig. 36; A, B eruptions as first seen; C, D last traces.)

On calling at Kew Observatory a day or two afterwards, Carrington learned that at the very moment when he had observed this pheno-menon the three magnetic elements at Kew were simultaneously dis-turbed. . . . This disturbance occurred as nearly as possible at 11^h 15^m G.M.T., on September 1, 1859, affecting all the elements simultaneously,

and commencing quite abruptly. Carrington's observation was confirmed by Hodgson.

We should now be inclined to regard this simultaneous disturbance as due to a fleeting excess of ultra-violet light, and the great storm beginning about 18 hours later, at 4.50 a.m., September 2, as the effect of the sun's corpuscular radiation. Balfour Stewart was assured by Sabine that, 'for excessive violence of character and length of duration, this storm had never been surpassed by any similar phenomenon, which has occurred in his long and varied experience'

This storm, especially the remarkable coincidence of the solar flare with the beginning of the short bay-disturbance, has been often discussed. Lord Kelvin regarded the evidence as a mere coincidence. We quote from a paper [28 a] by Ellis (1901):

'Carrington, one of the observers of the solar movement, while considering the phenomenon as deserving of notice, said that "he would not have it supposed that he even leans towards hastily connecting them". An occurrence so striking attracted attention, but unfortunately the narrative became repeated with exaggeration of statement, inducing a belief in direct connexion. But time at last showed that here was apparently misconception, for although the Sun, in the ordinary routine of solar work, has since been unremittingly watched, and a continuous photographic magnetic record also maintained, similar conditions have not been again observed. The magnetic motion, in the case in question, was in itself in no way remarkable, was indeed slight, . . . and much greater movements are also sufficiently numerous, but yet direct correspondence has not been made out.'

If the clear relation found in the years 1935 and 1936, between solar eruptions and radio fade-outs (10.3), had not given additional weight to the coincidence of Carrington's observation with small but significant simultaneous magnetic disturbances, geophysicists would even to-day maintain the cautious standpoint expressed by Ellis. But Carrington's solar eruption, seen in the total light (not in one spectral line or band), remains a unique observation.

9.24. Simultaneity of commencement of magnetic storms.
Figs. 1 and 37 illustrate the fact that great magnetic storms often commence almost simultaneously over the whole earth. Many investigations have been made to determine the degree of this simultaneity, namely, whether the times of commencement at different observatories differ by minutes, seconds, or by still smaller time-intervals. Such studies are limited by the degree of accuracy of the time-keeping and time-marking (on the magnetograms) at different observatories, by the quickness of response of the recording magnets to changes in the earth's

field, and by the rapidity and definiteness of these changes themselves. Obviously, also, such studies involve much labour, both in making the actual time-measurements of commencements at many different observatories, and in collecting and comparing such measurements.

Adams [37] and Ellis [38] showed in 1880 that the commencements may be simultaneous within a few minutes. In 1910 Bauer [29–31]
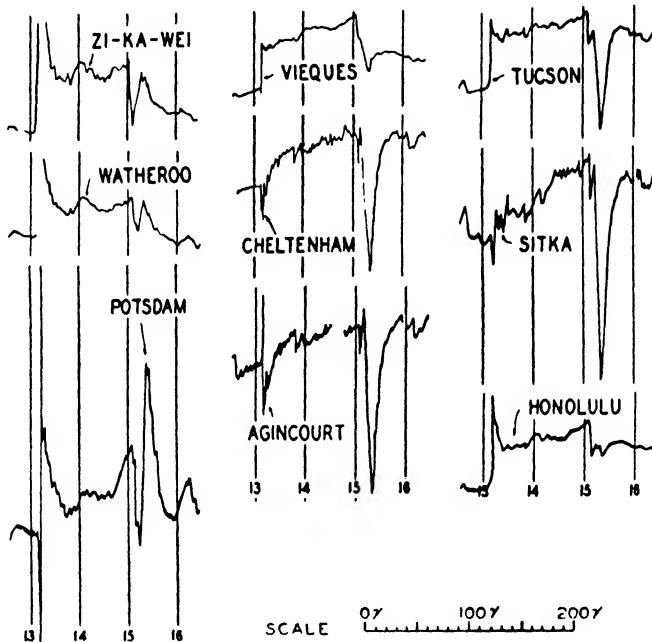
FIG. 37. Sudden commencement in horizontal intensity at various observatories, 1921, May 13. (After Peters and Green [2.83]. Greenwich Mean Time)

revived interest in the question and collected data for the initial epochs of many storms, from which he inferred that the time of propagation 'round' the earth was of the order of 3 or 4 minutes; he indicated the supposed direction and speed of propagation for several particular storms. These conclusions were contested by various other writers, and in order to arrive at more agreement Bauer collected new data for fifteen magnetic storms, from thirty-two observatories. This compilation, and the further studies of the matter that have since followed, incidentally increased the interest and care taken in the accurate time-scaling of magnetograms.

Bauer's new data were discussed by Angenheister [33], Chree [34], and Chapman [14], who found in them no support for Bauer's original

conclusions, but inferred that the available set of time-differences between the commencements at the various stations did not constitute a safe basis for the determination of disturbance-velocities, but rather gave an indication of the relative accuracy of the observations.

Chree remarked: 'As regards these "sudden" changes three things are conceivable: they may be absolutely simultaneous at different stations; there may be a very small difference of time, corresponding to the rate of propagation of electromagnetic waves; or, finally, there may be, as Dr. Bauer concludes, longer intervals, amounting to several minutes, for stations remote from one another. Under existing conditions of registration one cannot decide between the first two possibilities. As between these two and the third, a decision should not be impossible.'

Chapman suggested that there is at least another possibility, namely, that the time-differences may be of the same order as the time taken by a stream of corpuscles from the sun to sweep across the earth (24.3); this time is about 30 seconds; he also urged that the most appropriate way of arranging the time-differences for a series of storms is by reference to local time and latitude, instead of by reference to Greenwich longitude and latitude, as had hitherto been done; this was because of the likelihood that the cause of the storms is some agency proceeding from the sun, whose effect upon the earth, at particular observatories, will depend upon their position upon the earth, at the time, relative to the sun.

Later observational studies of this question have narrowed down the time-intervals to something of the order of a minute, but as yet we have no certain knowledge of the intervals to within a few seconds, or of their dependence upon the position of the observatories at the time when the storms begin. La Cour [36] found practical simultaneity (within the error of 3 seconds in the determination) for the beginning of the storm 1933 April 30 at the stations Godhavn, Julianehaab (Greenland), and Copenhagen, and the time there was only a few seconds different from the times for Huancayo (Peru) and Watheroo (Australia).

The initial impulse of sudden commencements has been discussed in § 11.

BAYS, PULSATIONS, AND MINOR DISTURBANCES

10.1. Bays. On days which are magnetically almost quiet, the magnetic records sometimes show small disturbances of a simple character. In the horizontal intensity H the curve is deflected from the normal course of the daily variation, the deviation gradually increasing till H has attained a maximum or minimum value; then gradually or with small oscillations H returns to its undisturbed value, without any appreciable after-effect. Because of the resemblance of this deviation of the magnetic curve to an indentation of a coast-line on a geographical map, these disturbances are called *bays* (Chree [G 15]), positive or negative according as H increases or decreases. They last for an hour or two. In the vertical intensity Z their shape is similar to that in H; the D curve is usually less disturbed.

These bays can be studied by drawing, on the magnetic record, a curve representing the continuance of the undisturbed curve 'across' the bay, and measuring the 'disturbing force' causing the bay, from this interpolated curve. The discussion of these bays may well throw light on similar variations observed during magnetic storms (cf. 9.17–18); in fact, many of the irregular features of a magnetic storm may be described as due to the superposition, on the general planetary phenomena described in 9.1–16, of a series of bay-shaped disturbances of varying sign and of a more regional character (9.17).

Steiner [1] studied the bays recorded on the magnetograms of O'Gyalla Observatory near Budapest, during the years 1906–17. He collected and measured 428 positive and 103 negative bays, about 40 and 10 per year. During the months May to August there were on the average about 2·7 bays per month, whereas the average number for the remaining part of the year was 4·6 per month. The positive bays showed a marked tendency to occur most often during the eight night hours centred at midnight; no positive bays were observed in the daytime between 8^h and 16^h. The (rarer) negative bays, on the contrary, occurred most often during the early afternoon hours. By dividing the bays into groups according to the intensity of the disturbance in H, it appeared that negative values of ΔZ are associated with positive values of ΔH, and vice versa, as Table 1 shows.

These numbers indicate that the disturbing vector at O'Gyalla is directed northward, at an upward inclination of 5° to 15° to the

TABLE 1. *Average maximum disturbing forces (γ) in H and Z, O'Gyalla*

	Positive bays				Negative bays		
Number of bays . .	18	62	129	77	18	34	20
ΔH	$+18{\cdot}6$	$+12{\cdot}0$	$+7{\cdot}5$	$+3{\cdot}8$	$-3{\cdot}4$	$-7{\cdot}4$	$-13{\cdot}0$
ΔZ	$-2{\cdot}7$	$-1{\cdot}8$	$-1{\cdot}3$	$-1{\cdot}1$	$+1{\cdot}3$	$+1{\cdot}1$	$+2{\cdot}0$

horizontal, during a positive bay, and in the opposite direction during a negative bay.

The loop described by the end-point of the magnetic field vector
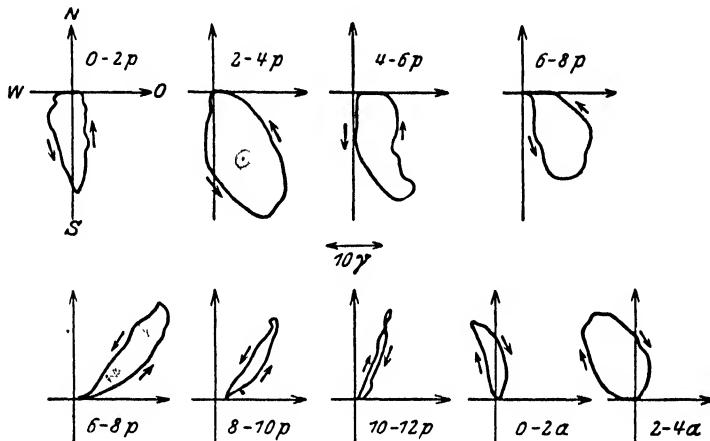
FIG. 1. Horizontal projections of the loops described by the end of the magnetic field vector during 'bays' at O'Gyalla at different times of the day. Upper row, negative bays; lower row, positive bays. (After L. Steiner)

during a bay is usually almost a plane curve (McNish [3]). The horizontal projections of these loops, averaged by Steiner for the bays occurring at various times of the day, are given in Fig. 1.

Similar statistical studies have been made by Wiechert [4] for Gross-Raum (near Königsberg, Germany), 1929–33, and by Lubiger [5] for Samoa (1914–20). They also found that positive bays are more frequent than negative bays, the ratio being about 2 to 1; this is less than the ratio (4 : 1) found for O'Gyalla. The daily change of frequency is similar at Gross-Raum to that at O'Gyalla. At Samoa, the night maximum of frequency for positive bays is even more pronounced than at O'Gyalla, and the negative bays have a pronounced noon-maximum. The *annual* frequency-variation of positive bays at Samoa shows a regular double wave with maxima near the equinoxes.

McNish [3] has studied several bays observed during the Second International Polar Year; he computed equivalent overhead currents

in the manner described in 7.7, as if the disturbances were due wholly to such currents (though, of course, he recognized that during such comparatively rapid variations induced earth currents contributing materially to the disturbance would certainly be present). He found, as Birkeland and Chree [G 15, p. 118] had likewise found in earlier investigations, that bays in middle latitudes are often accompaniments
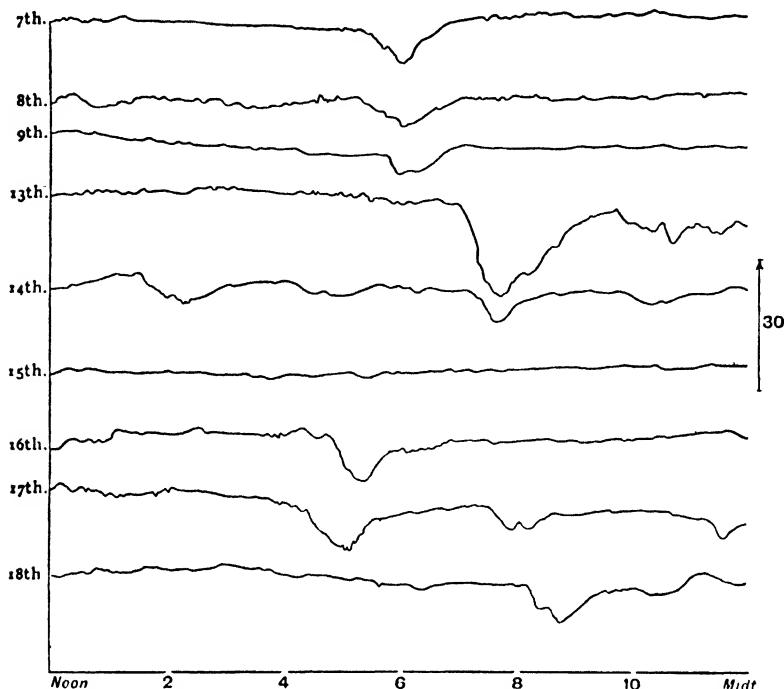
FIG. 2. 'Bays' in the Kew declination curves, showing some apparent repetitions on successive days, in February 1911. (After C. Chree)

of Birkeland's polar elementary storms. The specially intense disturbance in polar regions is due to a westward current concentrated along the night side of the auroral zone. In some cases the flow seems to be continuous around the zone, but usually there is an eastward current in part of the zone, opposite to the maximum westward current in the zone. The current-flow in the circuit is mainly westward along the night portion of the northern auroral zone, southward along the sunset meridian, eastward in the tropics (thus producing a *positive* bay in middle latitudes during the night hours), and northward along the dawn circle.

10.2. Repetitions on consecutive days. Any one looking through a long series of photographic magnetic records can hardly fail to notice

a certain peculiarity to which Capello, of Lisbon, seems first to have drawn attention, and which Balfour Stewart discussed in 1882 in his famous Encyclopaedia article [G 7] on terrestrial magnetism. This peculiarity is the repetition of some characteristic disturbance of the curve, once or more, on successive days, mostly with intervening intervals of about 24 hours. In spite of slight changes, especially in the amplitude of the repeated oscillations, the similarity is so pronounced as to suggest a physical relationship between these features. A good example is furnished by the 'precursors' observed at Potsdam before a disturbance of 1924 January 29, which was accompanied by aurora. During the aurora, the east declination increased by about 48 minutes of arc within 15 minutes of time, and decreased afterwards to its previous value. This bay-shaped feature appeared very distinctly, though with smaller amplitude, at the same hour on the two preceding days; the amplitudes of the two earlier bays were 2 and 4 minutes.

This phenomenon well deserves detailed investigation. In this connexion, the peculiar magnetic disturbances of 1937 April 24–8 may be mentioned, which, according to Fanselau's description [9.26], were interrupted by quiet intervals of several hours, and began, with sudden commencements, a few hours later on each successive day (Fig. 3).

10.3. Magnetic effects connected with solar chromospheric eruptions and radio absorption fadings. During observations with the spectrohelioscope [6] at the Huancayo Magnetic Observatory, on April 8, 1936, there occurred the most spectacular sunspot eruption, as regards extent and intensity, that had been witnessed during a year's observations there. It began at $16^h 45^m$ and lasted till $17^h 03^m$ G.M.T. Such spectrohelioscopic observations naturally extend over only a small fraction of each day.

At Mount Wilson observatory no unusual activity was noted until the morning of April 8. A calcium spectroheliogram taken at $16^h 22^m$ G.M.T. was quite normal. Four hydrogen spectroheliograms, beginning at $16^h 47^m$, revealed that an exceptionally brilliant eruption over a large spot-group in longitude 5° west and latitude 20° north had started a few minutes before; the bright flocculi reached their maximum at $16^h 52^m$, and were much weaker at $17^h 05^m$; at $18^h 30^m$ the appearance of the sun's disk was much the same as before the eruption.

The magnetogram at Huancayo recorded a conspicuous sudden commencement at the same time; H increased by 108γ from $16^h 46^m$ to $16^h 51^m$, then it decreased by 139γ from $16^h 51^m$ to $17^h 28^m$. Thereafter H was quiet and undisturbed, the H variation being definitely of zero
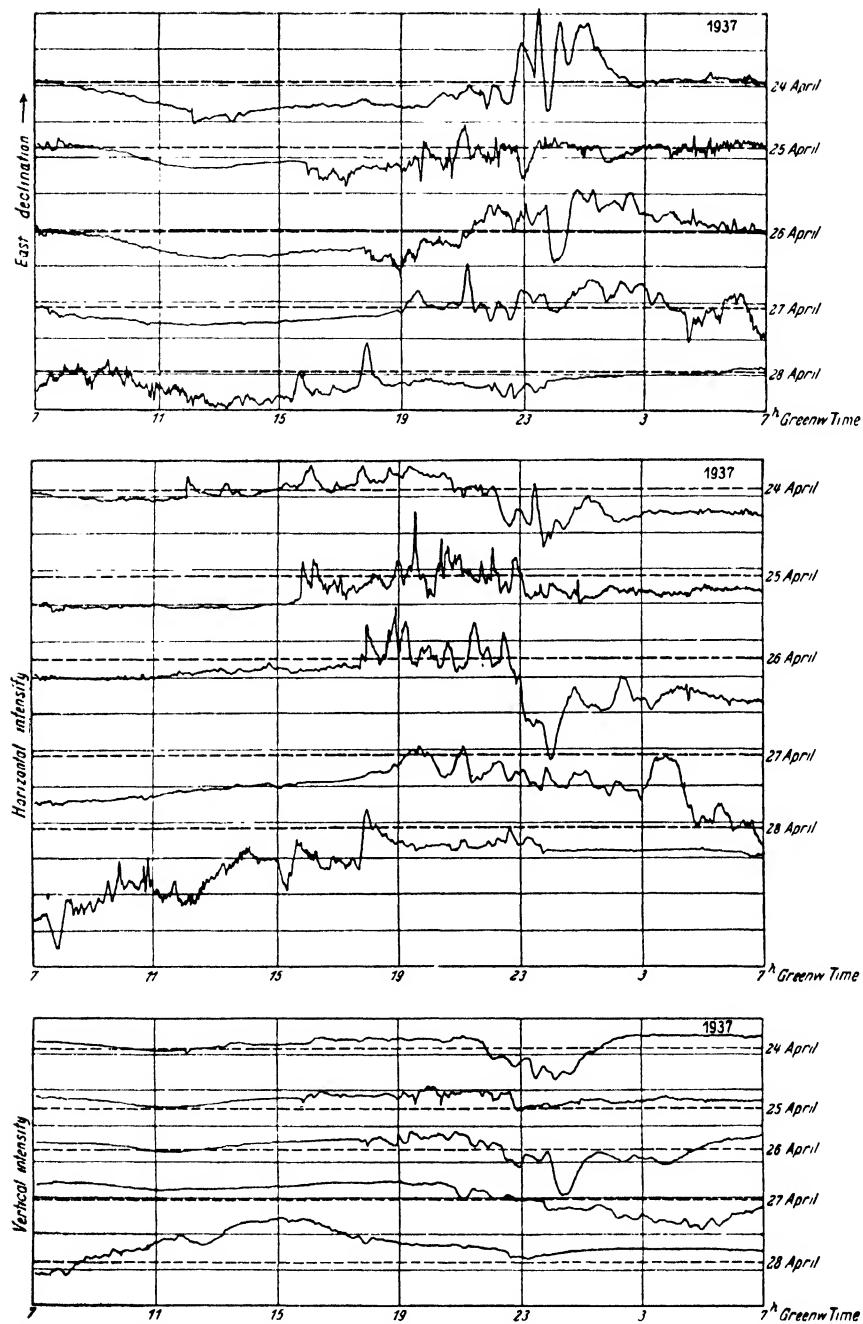
FIG. 3. The magnetic disturbances of April 24 to 28, 1937, as registered at Niemegk. The full horizontal lines are drawn at distances corresponding to 100γ; the undisturbed levels are indicated by the broken lines. (After G. Fanselau)

PLATE 25
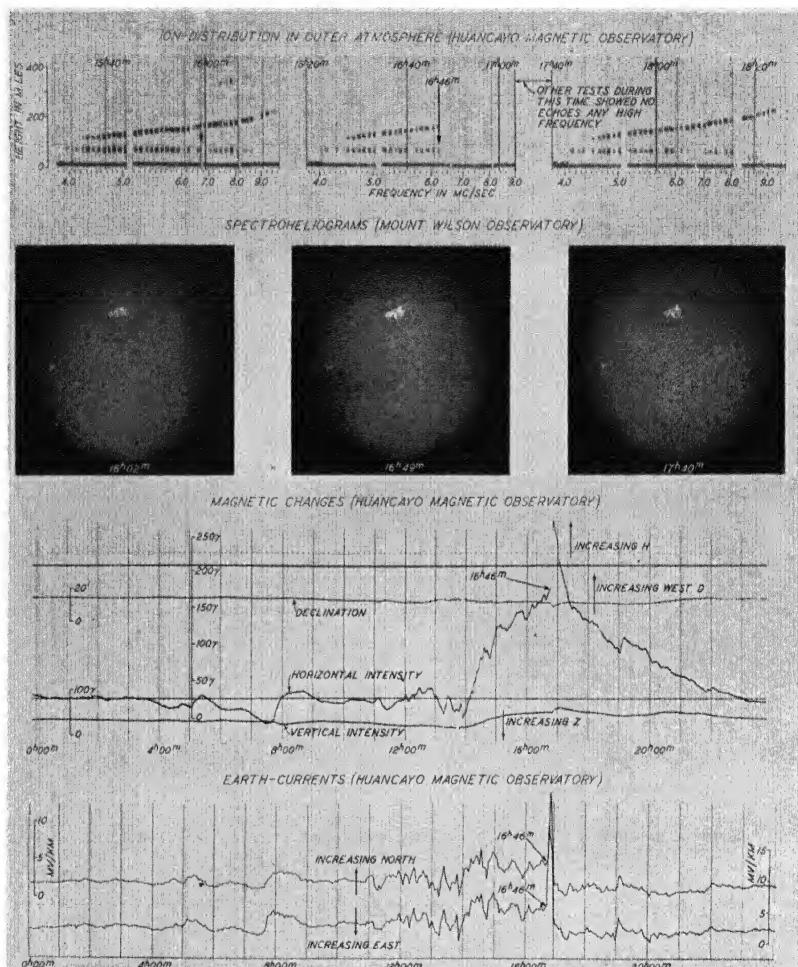
Magnetic, radio, and earth-current disturbanc esassociated with brilliant solar eruption
1936 April 8: Greenwich mean time throughout (Carnegie Institution of Washington)

magnetic character; D increased by 2·1′ from 16^h 47^m to 16^h 51^m; Z increased by 1·1 γ from 16^h 47^m to 16^h 49^m. This entire disturbance was thus confined to the short period of about 2 hours. The magnetic traces for the preceding and following days were quite undisturbed.

The earth-currents at Huancayo were varying smoothly until 16^h 45^m; at that minute all four lines showed radical changes in the earth-potentials, by about 30 millivolts. By 17^h 03^m the potential had returned to normal. The multifrequency ionospheric record at Huancayo from 16^h 20^m to 16^h 44^m covered a range of frequencies from 3,800 to 8,600 kilocycles; reflections from the ionosphere were obtained consistently. From 16^h 45^m no reflections were obtained from any layer. Another record beginning at 17^h 40^m gave nothing unusual. Plate 25 shows the Huancayo magnetic, radio, and earth-current records associated with this solar eruption of 1936 April 8.

Similar cases of simultaneous solar eruptions or *flares*, short magnetic disturbances (Fig. 4), and radio fade-outs, have been observed in increasing number since 1935, when attention was called to this association by Jouaust [7 a] and Dellinger [7]. A record of ionospheric reflections during such a solar flare is discussed also in 15.5, Fig. 15.1.

AUG. 25, 1936
HUANCAYO, PERU

Fig. 4. Magnetic disturbance at Huancayo during the solar flare and radio fade-out of 1936 August 25. (After J. A. Fleming [9])

Dellinger [7] collected data for more than 100 radio fade-outs in the years 1935 and 1936 (Figs. 5–9), and the material is rapidly increasing. Many of these were accompanied by magnetic effects similar to that of 1936 April 8. They share the characteristics of the radio effects, i.e. simultaneity throughout the portion of the earth affected, absence
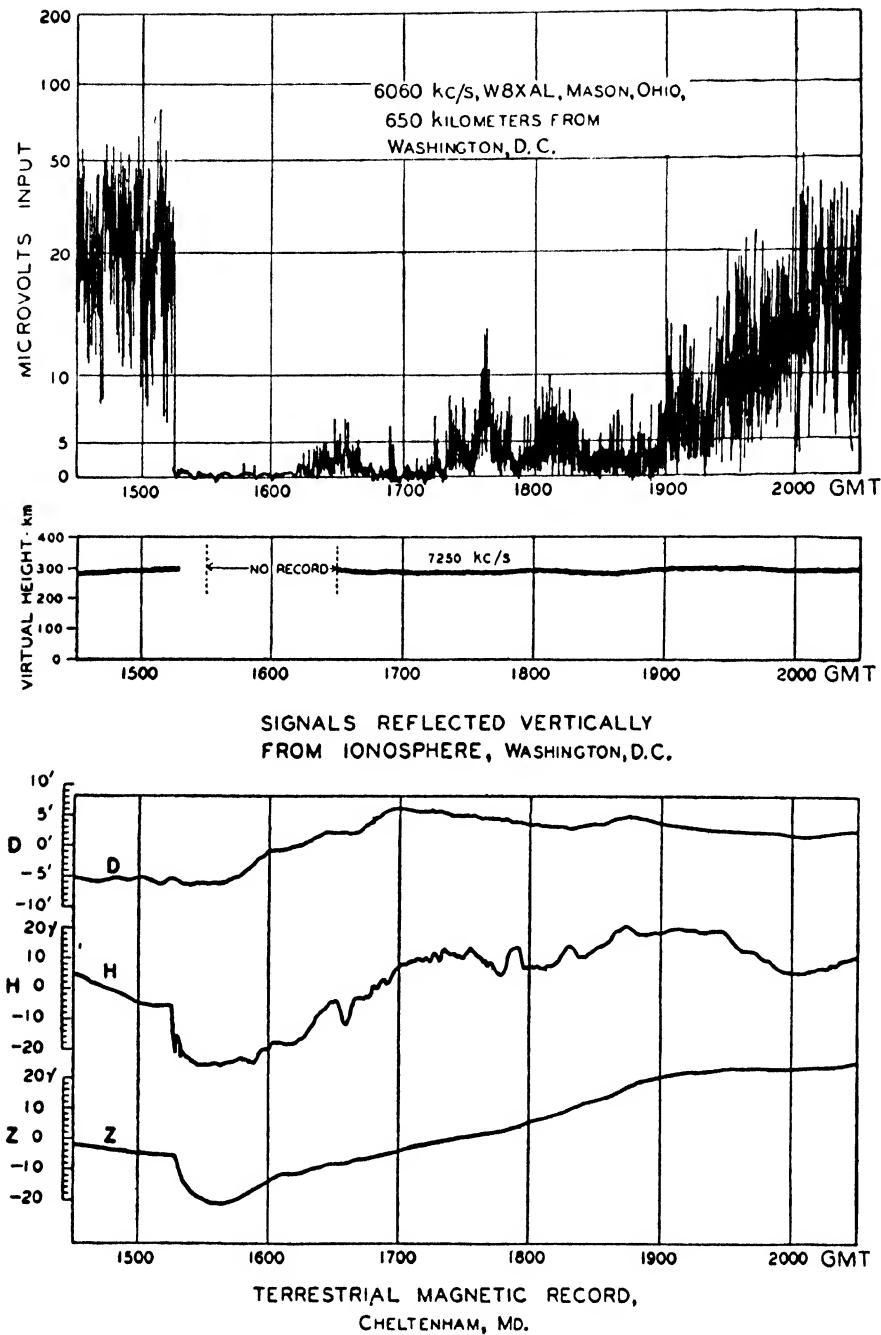
FIG. 5. The radio fade-out (low signal intensity) and magnetic disturbance on February 14, 1936. (After J. H. Dellinger)
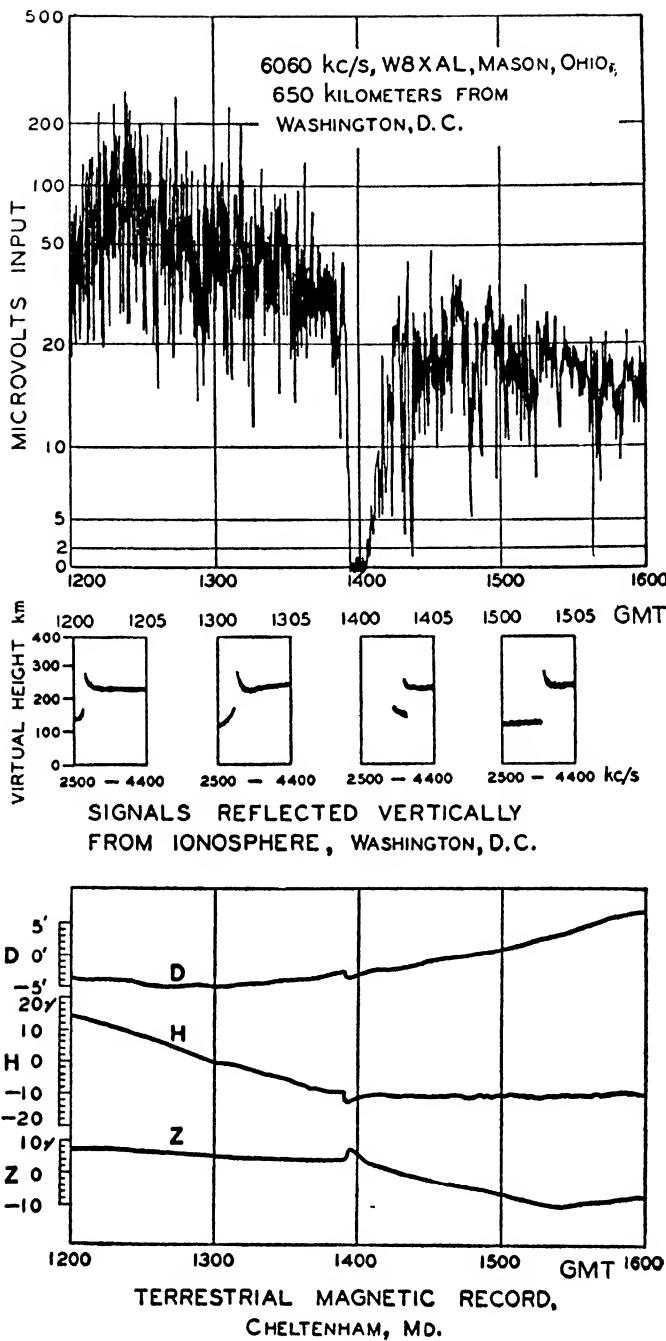
FIG. 6. The radio fade-out (low signal intensity) and magnetic disturbance on April 6, 1936. (After J. H. Dellinger)
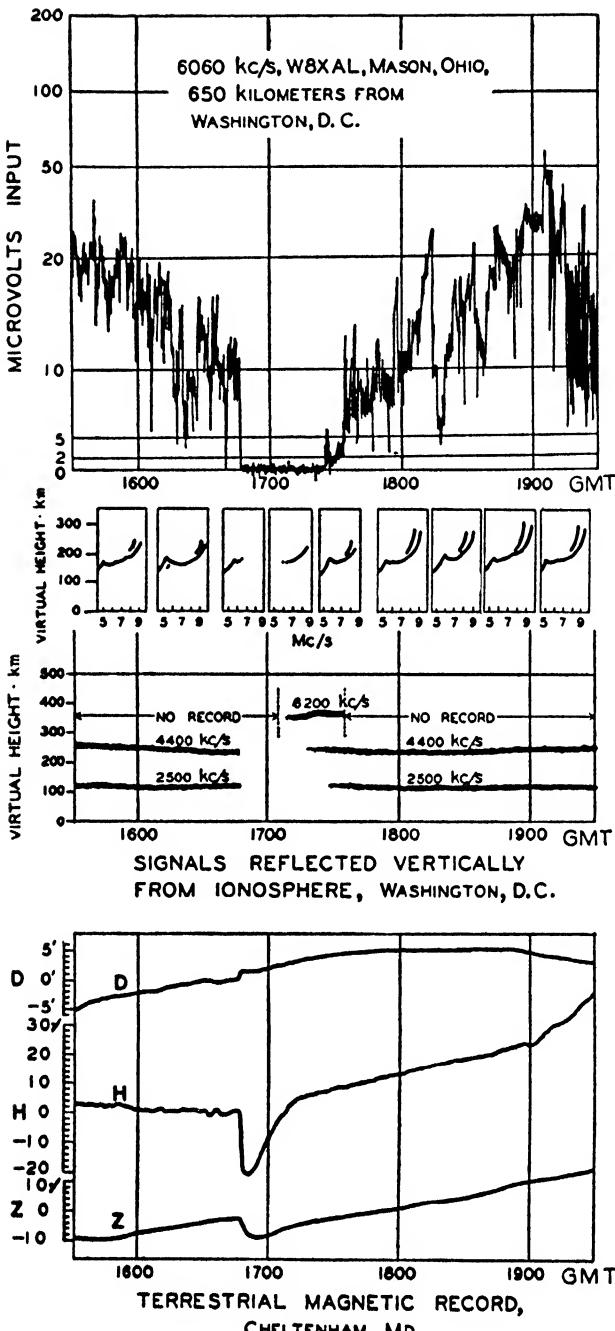
FIG. 7. The radio fade-out (low signal intensity) and magnetic
disturbance on April 8, 1936. (After J. H. Dellinger)
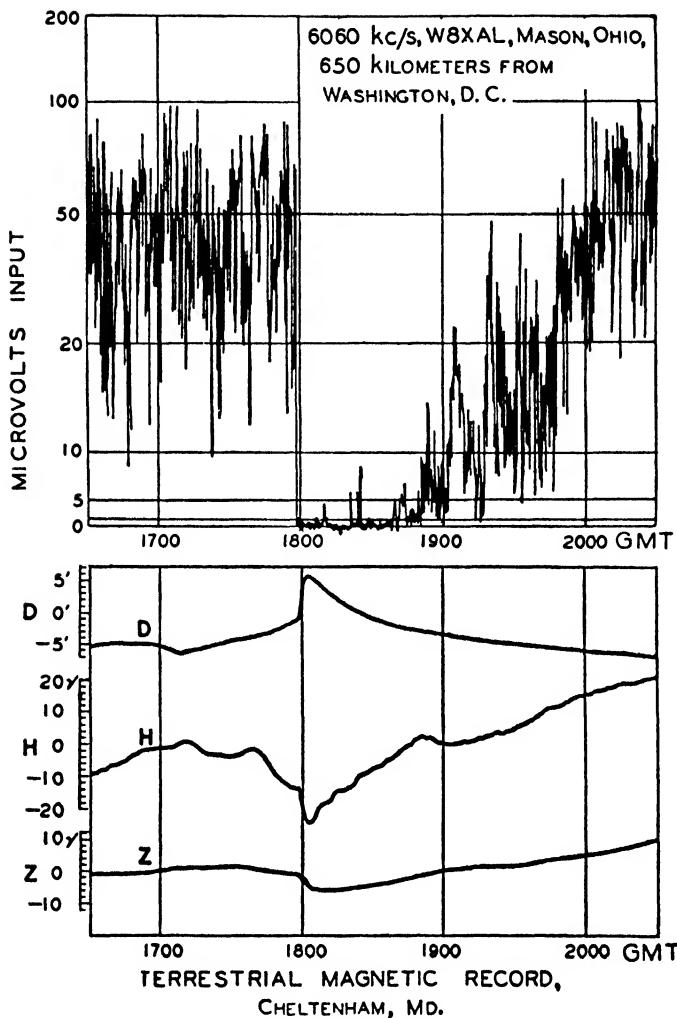
FIG. 8. The radio fade-out (low signal intensity) and magnetic
disturbance on May 28, 1936. (After J. H. Dellinger)
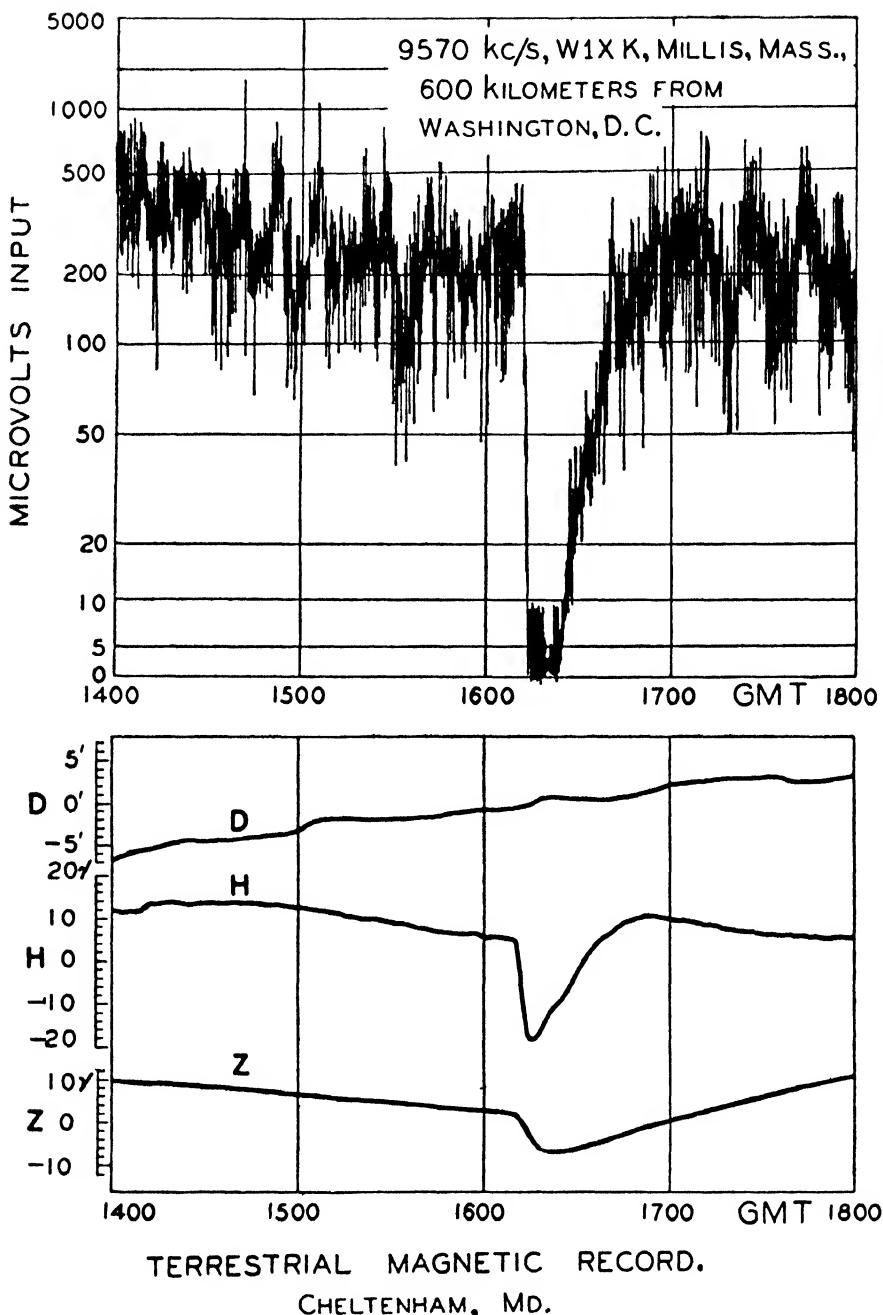
FIG. 9. The radio fade-out (low signal intensity) and magnetic disturbance on
November 26, 1936. (After J. H. Dellinger)

of disturbance over the dark hemisphere, suddenness, and maximum intensity in the regions where the sun's radiation descends vertically. But in the Greenwich records, Newton and Barton [5.12] found magnetic effects of the order of 10γ in only three definite cases out of twenty-nine cases of associated radio-fadings and bright eruptions.

Fleming [9] and McNish [10] showed that the magnetic effect is an augmentation, over the sunlit hemisphere, of the normal daily variation, supposedly due to increased atmospheric ionization, by ultra-violet light from the solar eruption, at the base of or below the E-layer (Fig. 10); this magnetic effect may therefore be referred to as an S_q-augmentation.

The classical case of the solar eruption of September 1, 1859, as observed by Carrington, has been discussed in 9.23.

Some of the magnetic effects have a resemblance to 'bays'. Although the study of the magnetic effects connected with radio fade-outs has only begun, it seems that ordinary bays are a different phenomenon, especially since they occur frequently during the night hours.

10.4. Micropulsations observed with ordinary magnetograms. In a description of the 'great magnetic disturbance which extended from August 28 to September 7, 1859, as recorded by photography at the Kew Observatory', Balfour Stewart [12] says:

'From the serrated appearance of the curves, it will be seen that a force tending to increase one of the elements was generally followed after a short interval by one of the opposite description, and vice versa. The exertion of the disturbing force was thus of a throbbing or pulsatory character. The interval of time between two of these minute pulsations may be said to have varied from half a minute, or the smallest observable portion of time, up to four or five minutes.—This pulsatory character of the disturbing force agrees well with the nature of its action on telegraphic wires, in which observers have noticed that the polarity of the current changes very frequently.'

A number of special studies [13–22] have since then been made of these 'pulsations' (as van Bemmelen [13] called them) or 'elementary waves' (Eschenhagen [14]). They occur not only during storms, but also on otherwise quiet days. The periods, measured from crest to crest, lie mostly between fractions of a minute and 3 minutes; they appear as serrations on the ordinary magnetograms, where the time-scale is about 3 minutes for 1 mm., and as regular waves on the quick-run magnetograms, with a time-scale of about 12 seconds for 1 mm. Their regularity sometimes suggests that they might be due to resonance effects within the instruments; but this is ruled out by the fact that
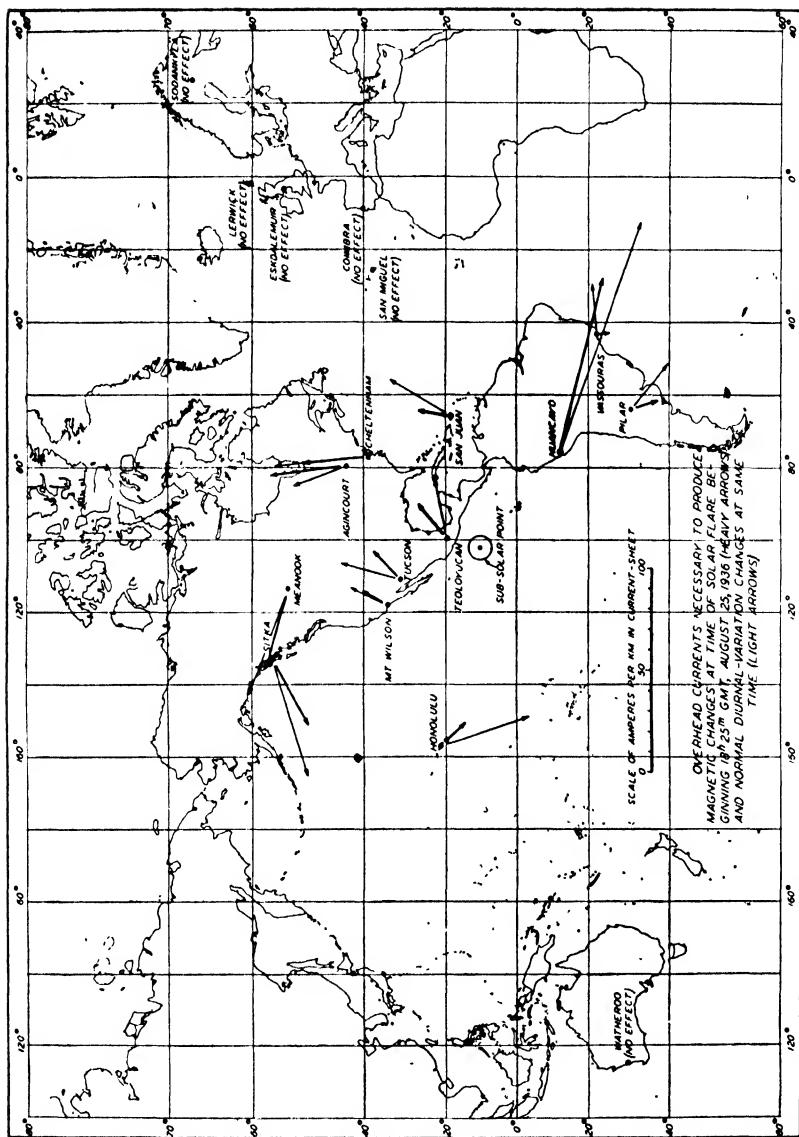
Fig. 10. Diagram illustrating the restriction of the magnetic changes, during a solar flare, to the sunlit hemisphere, and the similarity of the superposed disturbing electric current system to the existing S current system.
(After A. G. McNish)

the magnets in the instruments have much shorter free oscillation periods, 10 seconds down to 1 second, resulting in an indistinct burr on the trace when the magnets are set swinging. The amplitudes may reach several γ; in rare cases, giant (or Rolf) micropulsations have been locally observed in Northern Scandinavia (Abisko, Tromsö [19, 22]) with ranges of more than 30γ; Leiv Harang [19] has called them 'sinusoidal oscillations' because of their very regular form (Fig. 11).
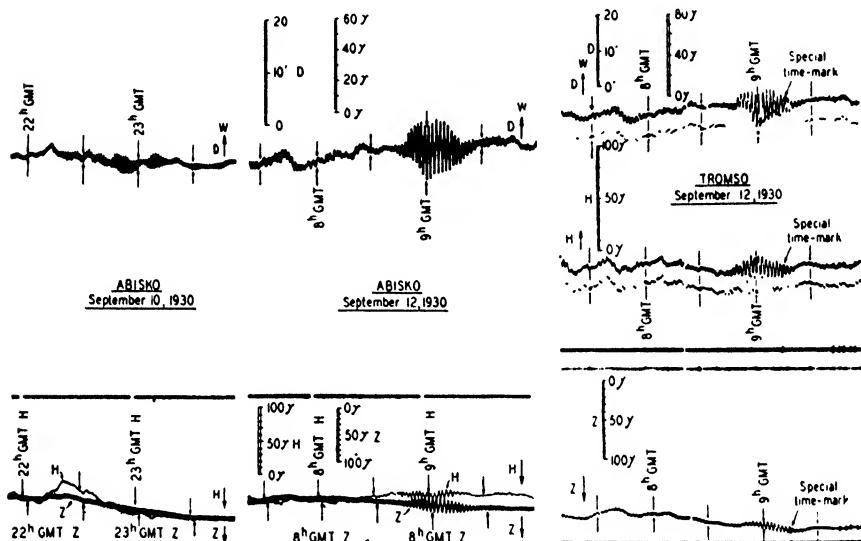
FIG. 11. Giant micropulsations recorded at Abisko and Tromsö, September 1930. (After B. Rolf)

The pulsations of 1 to 3 minutes period occur most frequently around midnight. Lubiger [18] has studied the pulsations at Samoa, and has constructed frequency tables relating to them. From these it appears that the frequency of pulsations during the 4 hours centred at midnight is higher than that during the 4 hours at noon, in the following ratios: Samoa 4·2, Batavia 4·3, Potsdam 110, Zikawei 3·4; at Tromsö, 1929–36, oscillations are most frequent during the hours 2 to 4, and are rare in the day-time. Lubiger found that at Samoa the shorter periods (under 90 sec.) showed the night maximum more sharply than the longer periods. The distribution over the seasons is fairly uniform; it seems uncertain whether the pulsations are more frequent in disturbed or in quiet years of the 11-year cycle.

At Huancayo, Peru, irregular pulsations of quite appreciable amplitudes (several gammas), with periods of 1 minute or more, are a con-

spicuous feature of the records for the horizontal intensity in the day-time; the photographic records have a jagged appearance even on magnetically quiet days.

10.5. Pulsations on quick-run magnetograms. For the study of pulsations, Eschenhagen [14] (1896) introduced 'Feinregistrierungen', that is, magnetographs with a wider time-scale than usual (nearly 24 cm. for an hour) and of high sensitivity (0.3γ per mm.). He found pulsations of 30 seconds period in H. The introduction of the Schmidt-La Cour type of quick-run recorders, with a time-scale of 18 cm. for an hour (1 mm. for 20 sec.), has made possible continuous quick recording.

Sustained oscillations of shorter periods (between 20 and 80 sec.) were examined in the quick-run Watheroo records for 1933 by Johnston [21]. On 69 days, pulsations of very short periods (between 20 and 30 sec.) were found, and on 114 days, pulsations of period between 30 and 80 seconds. Of the 69 days with pulsations of very short periods, between 20 and 30 seconds, only one was quiet (international character-figures 0·0 to 0·3), while of the 114 other days, 44 were quiet.

Van Bemmelen [13] studied the micropulsations recorded at Batavia and at an auxiliary station 800 km. away, and found simultaneous pulsations. At Batavia he never found natural pulsations of the vertical intensity, although artificial pulsations of 30 seconds period and 1γ amplitude (produced by moving an adjacent magnet) were faithfully recorded by his variometer. In higher latitudes the vertical intensity is affected by pulsations, but generally with smaller amplitudes than the horizontal components.

Angenheister [16] found examples of pulsations recorded simultaneously at stations so widely separated as Cheltenham (near Washington), Samoa, and Tsintau, and concluded that pulsations occur simultaneously over the whole earth. This, however, is not confirmed by later observations of pulsations in the polar regions; the giant micropulsations recorded at Abisko [22], for instance, could only be detected in the records of stations not more than 1,000 km. distant.

During the Polar Year 1932/3, no pulsations of the type described by Rolf for Abisko, and by Harang for Tromsö (of about 1·5 minutes period), have been observed on the quick-run magnetograms at Thule (near the centre of the auroral zone) nor at Godhavn (magnetic latitude 70°). They seem, therefore, to be rare inside the auroral zone. But at the station in Iceland, Thorkelsson [23] found very pronounced and regular oscillations on several occasions a few hours after midnight; they seem therefore to occur near, but not far within, the auroral zone.

10.6. Rapid vibrations. Periods of 20 seconds (equivalent to 1 mm. on the adopted time-scale) are the limit of resolution for quick-run magnetograms. Considerably smaller periods, less than a second, have been observed with undamped magnets having free oscillation periods of the same order. The vibrations appear as a continuous broadening of the curves (reminiscent of a short shuttle) for D and H, with an amplitude of 20 to 30γ, while the heavier Z-magnet, with its longer oscillation period, remains mostly unaffected. They occurred at Tromsö on 42 days between September 1932 and June 1936, principally in the morning, with a frequency maximum between 8^h and 10^h. They occurred simultaneously at Bossekop and Tromsö, 170 km. apart, and are therefore not due to local artificial disturbances. Harang [19] considered it not impossible that mechanical vibrations of seismic origin (microseisms) might cause some of the recorded vibrations, but he excluded this explanation as a general one, because of the pronounced daily variation of frequency.

For Sodankyla 1932–5, Sucksdorff [20] describes these rapid micropulsations (with periods less than 2 or 3 seconds) as follows:

'Even a casual glance at the records reveals that rapid micropulsations occur in groups, and only from time to time. Usually they take place in the course of a few consecutive days, after which sometimes a week, or even a month, may pass before new oscillations occur. It is seldom that such an oscillation occurs singly; usually there are many oscillations, interrupted by short intervals, for one hour, or for several hours, or sometimes even for more than 24 hours continuously, when the curve often resembles a pearl necklace consisting of oval pearls of different sizes.'

About one-thirtieth of all hourly intervals in the four years were affected by such rapid vibrations; they are more frequent in the day-time than at night.

Sucksdorff mentions that lightning striking the ground near the observatory, or other irrelevant causes deflecting the magnet, produce quite different vibrations; these damped free oscillations of the magnet give wedge-shaped markings, easily distinguished from the shuttle-type described above.

10.7. Pulsations recorded by induction and otherwise. Rapid oscillations of any variable quantity are magnified in its time-derivative. The micropulsations are therefore likely to be found in the records of the 'induction-loops' described in 2.11, and have been observed on several occasions, with different instrumental arrangements, in dZ/dt by Ebert. Rössiger measured such pulsations in Z

directly. The periods found were of the same order as those found on the magnetograms.

So far, no systematic attempt has been made to bridge the gap between these natural magnetic pulsations with periods down to 1 second, and the natural electromagnetic high-frequency waves (with periods up to 10^{-4} sec.) observed in 'atmospherics' or 'statics' by wireless technique. Of course this region of the frequency spectrum, in civilized countries, is highly 'contaminated' by ordinary alternating current periods, which would make observation of natural effects rather difficult. However, the interesting relations interpreted by Schindelhauer [24], between the daily directional changes of the atmospheric disturbances of wireless waves and the lines of geomagnetic force, would seem to call for such studies.

10.8. Magnetic effects during solar eclipses. Many attempts have been made to discover whether a solar eclipse produces measurable changes in the geomagnetic field. Bauer [25–33] was specially active in the study of this question, and between 1900 and 1919 arranged many expeditions and joint programmes for special magnetic and allied observations in or near the track of totality and elsewhere, during eleven eclipse periods. The discussion of these and similar data is to be found in an extensive series of papers by Bauer, van Bemmelen [34], Chree [35], and others; views were divided as to whether or not effects reasonably attributable to the eclipses had been established. Bauer considered that they had been, whereas Chree was sceptical. Apparently the first attempt to find guidance from theoretical considerations as to the probable nature and magnitude of the effects that might be expected was made by Chapman (23.22); he concluded that eclipse effects of a specified nature might not be too small to be observable under favourable conditions (which he indicated). The subject seems to deserve further consideration, by re-examination of the data collected by Bauer (or in association with his efforts), and by new observations.

It should be added that, according to Chapman's theoretical discussion, partial eclipses (if, say, 70 per cent. or more of the sun is obscured at maximum phase) may be little inferior, in their magnetic effects, to total eclipses; thus totality in eclipses is of much less importance in geomagnetic than in astronomical studies, and the number of eclipses suitable for magnetic study is thereby enhanced. The frequent irregular changes in the geomagnetic field make it difficult to establish that small magnetic changes observed *during* an eclipse are *due to* the eclipse; hence a statistical treatment of many eclipses seems necessary.

MAGNETIC DISTURBANCE: STATISTICS AND SOLAR RELATIONSHIPS

11.1. Statistics in physics and geophysics. A description of a geophysical phenomenon which gives an account only of its *average* value or *average* variations (daily, seasonal, and so on) is not complete. Such an account must be supplemented by a discussion of the irregular variability of the phenomenon. Consider, for instance, the daily readings, at noon, of the air-temperature on summer days at a given station; the mere average is no sufficient abstract of the observations, because such information as, for instance, the frequency or intensity of hot and cold spells is needed for a complete picture.

In physics, such statistical questions are of importance in theoretical research on molecular or atomic phenomena. But in actual laboratory experiments they are usually rather unimportant, except in connexion with the theory of errors. This is the consequence of a well-known feature of laboratory experiments, namely, that the conditions of the experiment are made as simple and 'pure' as possible. If, for instance, one wishes to determine how the density s of a gas depends on the pressure p, experiments will be devised in which only p is changed, all the other conditions (temperature, etc.) being kept constant. In general, a function $u = f(x, y, z,...)$ is experimentally evaluated by changing only one variable at a time, etc.

But this simplification is only rarely possible in geophysics. Consider the magnetic field at a given station. It depends on the situation of the station on the earth, on the underground structure, and, above all, it is a very complicated function of the time; it depends on the year, the season, on solar and lunar time, and on the solar and magnetic activity at the time of observation and some time before. It is not possible for us to modify the course of these variables. In order to separate their influences, it is necessary to have data in which the variables appear in different combinations; we must also have observations from many stations distributed over the globe in order to find the influence of the geographical coordinates. In order to separate the various influences depending on the time, continuous registration over long series of years is of great importance. The discussions in 16.27 and 16.28 indicate how long the series of records should be, if, as is there shown, a full sunspot cycle of 11 years is to be regarded (in the study of variations depending

on the sunspot number), as equivalent to only twenty independent daily observations. Again, in connexion with the study of magnetic storms, it has to be remembered that since the beginning of photographic recording, only a few magnetic storms of extreme intensity have been observed (9.22).

Suitable statistical discussions, based on physical considerations, are of great value in elucidating the physical nature of geomagnetic phenomena, by making use of their natural variability. But although many elementary points connected with the statistical treatment of geophysical data may be found in any text-book on statistics [1–8], in many cases special methods must be developed, particularly in order to adapt the usual conceptions and methods to the case of time-series which have a marked conservation tendency (cf. 16.27–8).

11.2. The daily ranges of the disturbance-field in Greenwich declination. In this chapter various statistical questions concerning the incidence of magnetic disturbance, and the variation and distribution of its intensity, will be considered, but without discussing systematically how the disturbance-variation field D is distributed over the earth (as was done in Chapter IX), nor the relation to the solar rotation period (as in Chapter XII).

Some of the phenomena treated in this chapter can be investigated conveniently by means of the daily character figures (6.2), but, as already explained, these are unreliable for comparisons between different years or even between different seasons. For such purposes other measures of magnetic disturbance must be used, founded on some strictly quantitative basis.

The subjects here dealt with have been considered by many writers in different ways, using data from various stations. Some of the more recent studies, or those based on the largest amount of material, or those which seem to fit best into the general plan of this book, will be used for illustration. Similar studies based on other data will be mentioned in order to indicate how far the results quoted are known to be true for other regions.

The relative frequency of different degrees of disturbance will first be considered, using as material for the purpose the hourly values of magnetic declination at Greenwich throughout sixty-three years, 1848–63, 1868–1914 inclusive. The intervals taken as the unit periods in which the degree of disturbance is measured will be 25 solar hours, slightly in excess of the solar or lunar day; it is convenient, however, to refer to them as days; if 24-hourly periods had been treated by the

same method, closely similar results would have been found. (The results to be described were derived as by-products in the course of an investigation of the lunar daily variation of Greenwich declination [8.9].) The adopted measure of the disturbance on each day is the range, during the day, in the hourly differences obtained by subtracting the monthly mean solar daily inequality from the corresponding hourly values. This range will be denoted by R_D. It is the Greenwich declination measure of the field $\Delta_i F - S_q$ (6.9), except that the monthly mean

Frequency curves (%) of lunar days of different range (unit 1') in 3 seasons

Summer ————
Equinox — — — —
Winter ....
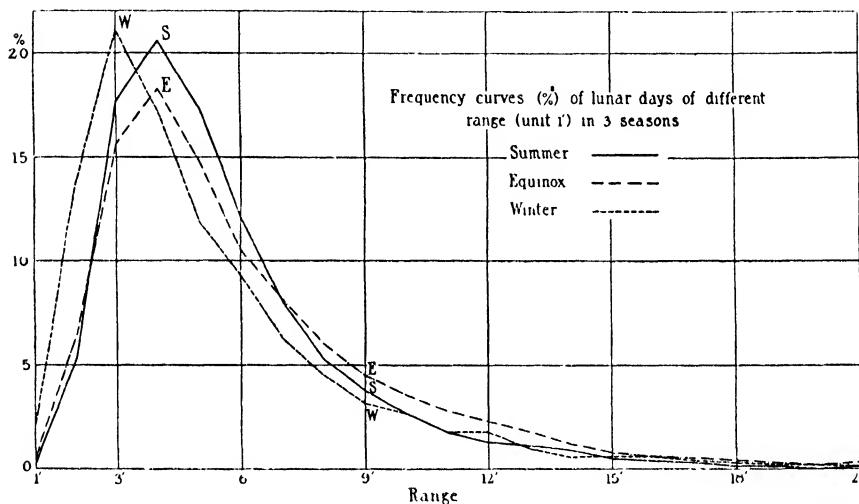
FIG. 1

daily S_q inequality from all days, instead of from quiet days (only), was used; the two inequalities are so nearly equal, in general, that the difference (for the present purpose) is immaterial. Neglecting this, and also the presence of the small lunar daily variation, R_D is a measure of the range, in Greenwich declination, of the D field defined in 6.9.

11.3. The frequency and averages of these ranges. The frequency of occurrence of days of different disturbance-range R_D will be considered for various groups of days. In each case the frequency-distribution will be expressed by percentages indicating what fraction of all the days in a group had ranges R_D lying within assigned limits. These were taken to be $(n \pm 0.5)'$, where n is an integer; the ranges may be expressed in terms of force-units at the rate of 5.2γ per $1'$.

Figs. 1, 2, and 3 indicate the frequency-distribution for various different groups of days (totalling 20,762). In Fig. 1 the days were grouped according to season, viz. (northern) summer, winter, and
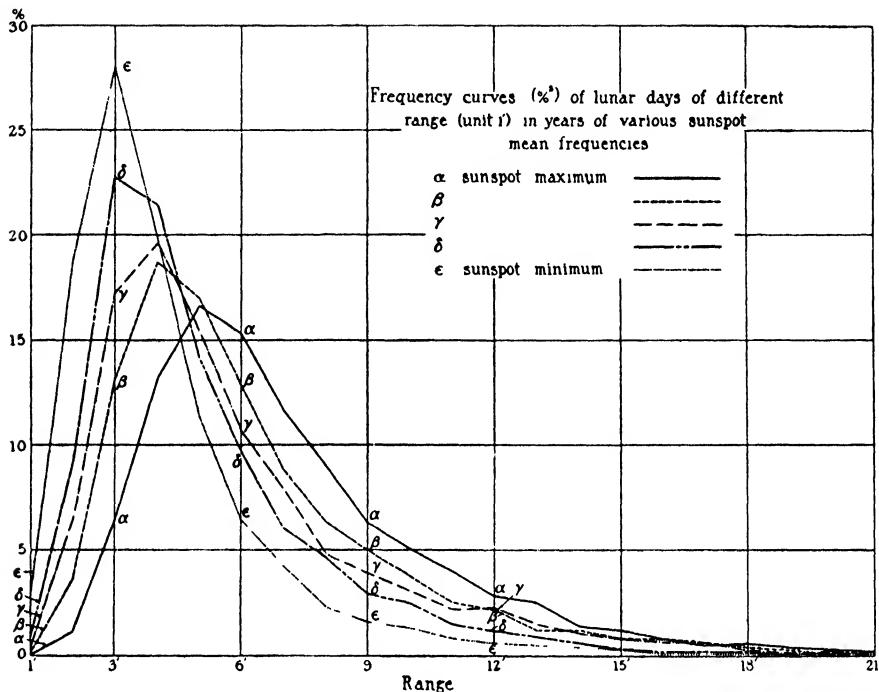
Fig. 2
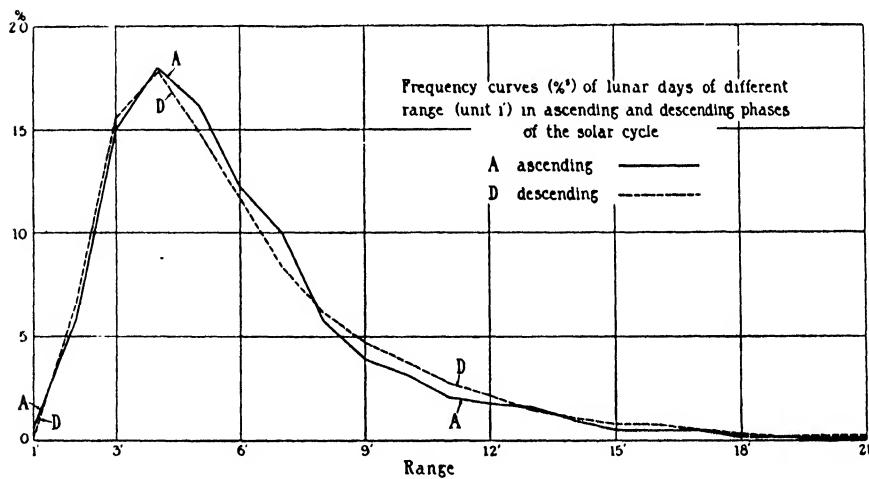
Fig. 3

*

equinox (p. 160). In Fig. 2 all the seasons were combined, but five groups of years were taken, denoted by α, β, γ, δ, ε, chosen according to their annual mean sunspot numbers; the number of years in a group varied from 11 to 14, and the mean sunspot number in each group is as follows:

Group .	.	.	.	.	α	β	γ	δ	ε
Mean sunspot number	.	.	96	63	44	19	6		

In Fig. 3, days from two groups (D, A) of 13 and 27 years respectively are included, taken from the descending and ascending phases of the sunspot cycle; the years were of medium sunspottedness, the mean sunspot numbers for the groups being practically equal, namely, 47·2 for D and 47·7 for A.

The frequency curves are of the same general form in all the cases, but Fig. 1 shows that, as between the three seasons, highly disturbed days are most frequent, and quiet days least frequent, at the equinoxes; quiet days are most frequent in winter. The most frequent ranges (corresponding to the peaks of the curves) are 4' in summer and at the equinoxes, and 3' in winter. The *mean* ranges are 6·10' at the equinoxes, 5·54' in summer, and 5·23' in winter.

Fig. 2 shows how greatly the distribution of disturbed days alters with the annual mean sunspottedness; the proportion of disturbed days increases, and that of quiet days decreases, as the sunspot number increases. The most frequent ranges in the groups α to ε, in order, are 5', 4', 4', 3', 3', and the *mean* ranges are 7·16', 6·28', 5·90', 5·11', and 4·18'. The percentages of days with a range exceeding 13' are 5·9, 4·8, 4·4, 2·1, and 1·3.

Fig. 3 shows that years of similar sunspottedness tend to be slightly more disturbed in the descending than in the ascending phase of the sunspot cycle; the percentage of days of range exceeding 13' is 4·4 in group D and 3·2 in group A, and the mean ranges are 6·15' in D and 5·84' in A. It has already been seen (5.5) that the sun's surface is not similarly disturbed at epochs of corresponding spot area in the ascending and descending phases of the sunspot cycle, the mean latitude of the spots being lower in the descending phase.

The seasonal variation in the degree of disturbance is further illustrated by the following mean values of R_D for the twelve calendar months, all years being combined:

Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
5·21'	5·74'	6·18'	5·99'	5·74'	5·28'	5·55'	5·60'	6·16'	6·08'	5·23'	4·74'

March and September have practically equal maximum ranges, while

June and December have minimum ranges, the December minimum being the lower.

Fig. 4, after Maurain [38], contrasts the annual mean sunspot num-
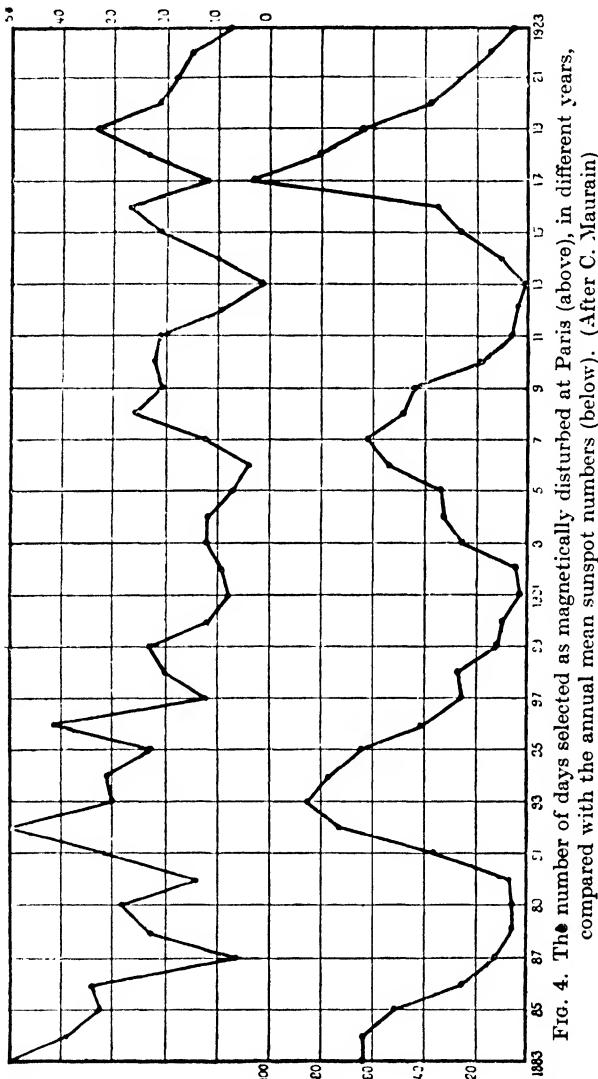
FIG. 4. The number of days selected as magnetically disturbed at Paris (above), in different years, compared with the annual mean sunspot numbers (below). (After C. Maurain).

bers (shown by the lower curve) with the number of days per year which were classed at Paris as disturbed. Eight hundred and fifty-five such days, or 57 per thousand, were selected in the forty-one years 1883–1923; their number per year varies between 50 (in 1883 and 1892) and 1 (in the sunspot-minimum year 1913). There is an obvious

correspondence in Fig. 4 between the sunspot variations and the variations in the number of disturbed days, but it is clearly much less close than for the average *intensity* of disturbance (Figs. 7 and 8).

11.4. The numerical magnetic character. The energy-density (per unit volume) of a magnetic field of total intensity F is $F^2/8\pi$. Bidlingmaier, seeking to define a physical measure A of magnetic activity for a certain time-interval, put A equal to the average value of $(\Delta F)^2$ during the interval, where ΔF is $F - F_0$, the vectorial difference between the intensity F and its mean value F_0 over the interval. Several observatories computed A according to his scheme for each day of the year 1915. Later, however, it was realized that the conception of A as a measure of the energy-density of the disturbing field is erroneous, because the variations of the energy-density are proportional to those of

$$F^2 = (F_0 + \Delta F)^2 = F_0^2 + 2F_0 \cdot \Delta F + \Delta F^2, \tag{1}$$

and the term $2F_0 \cdot \Delta F$ is usually about 1,000 times greater than the quadratic term ΔF^2 [25.13].

In rectangular components,

$$F_0 \cdot \Delta F = X_0 \Delta X + Y_0 \Delta Y + Z_0 \Delta Z. \tag{2}$$

Crichton Mitchell proposed, and the Association of Terrestrial Magnetism in 1930 decided [10], that as an experimental step, and for a period of years (terminated in 1939; see p. 213), international tables of numerical magnetic character figures should be prepared, giving

$$X_0 R_X + Y_0 R_Y + Z_0 R_Z \tag{3}$$

for observatories recording X, Y, and Z, and

$$H_0 R_H + Z_0 R_Z \tag{4}$$

for observatories recording D, H, and Z. These data refer to a Greenwich day; H_0, X_0, Y_0, Z_0 are the average values for the day, and R_H, R_X, R_Y, R_Z are the absolute ranges, that is, the differences between the highest and lowest values of X, Y, and Z respectively. The form of these measures, (3) and (4), is reminiscent of (2), but this similarity is deceptive, since the extreme values of X, which differ by R_X, occur generally at other times than the extreme values of Y and Z. The measures cannot therefore be regarded as having the special physical significance that might attach to figures proportional to the energy of disturbance.

Since 1931 daily values of the measures (3) and (4) have been compiled by van Dijk, for about thirty stations, and issued quarterly as 'Numerical Characters'. The tables also give $HR_H/10,000$, etc.,

individually (both H and R_H being measured in γ). They show a large seasonal change, being greatest in summer and least in winter; this is because the ranges, on quiet and ordinary days, mainly depend on the solar daily variation S_q, which is greater in summer than in winter [12–14]; in this respect these numbers cannot be considered as good indices of *disturbance*. Again, they show a year-to-year variation, their annual means increasing and decreasing in rough correspondence with the annual mean sunspot numbers. This is because the variations of the annual mean range of S_q, and of the annual mean intensity of disturbance (measured as in §§ 2, 3, independently of S_q), in the course of the sunspot cycle, are fairly similar. Here again the numerical characters fail in their purpose of giving an index of *disturbance*, and it seems clearly better to treat two independent variations separately, even if in some respects they vary in parallel with one another.

Other measures based on daily ranges have been proposed. For instance, Schmidt has given, for Potsdam since 1920, simply the sum $\alpha = (R_X + R_Y + R_Z)$ for each day; that is, the sum of the three sides of the rectangular space in which the movements of the end-point of the magnetic field vector (supposed to be plotted from a fixed origin) could be enclosed. This could be freed from the influence of the solar daily variation S by forming, from the monthly average S, the same expression α_S for the S, in order to compute the sum of the disturbance ranges α_D from $\alpha_D^2 = \alpha^2 - \alpha_S^2$.

A satisfactory objective measure of magnetic activity for single days now seems to be afforded (see p. 213) by the set of three-hourly range-indices K [44 a] for each day, although the international character figure C has not been superseded. For monthly and annual means of magnetic activity, the u-measure described in § 6 seems adequate.

11.5. The Tromsö measure of storminess, and similar measures of magnetic activity. As in § 2, the average solar daily variation has often been subtracted from the observed hourly values in order to obtain a measure of the 'disturbing force'; this was done, for instance, in the earlier volumes of Batavia observations. Again, the Results of the magnetic observations of the Auroral Observatory at Tromsö contain, in addition to the ordinary tables of hourly values, special tables of hourly values of 'storminess' for D, H, and Z, obtained in the following way. In the ordinary tables (A) of the hourly values, the quiet hourly values are marked, and re-entered on another set of tables (B) (not printed). The values thus entered usually appear in groups for a few quiet days, but sometimes only for intervals of a few

hours. From these 'normal' values monthly means for each hour are formed. With the aid of these means the many blank spaces in the tables B are filled up, thus giving a set of 'normal' hourly values for the disturbed periods also. The storminess S is then tabulated as the difference between corresponding entries in the two tables, A minus B. For each month the positive and negative values (PS and NS) are added, without regard to sign, to obtain the Absolute Storminess. Such a procedure is fairly satisfactory even for a polar observatory, where the quiet-day daily variation S_q is small, and the mean daily variation S, which is removed, may be composed partly of S_D (9.8), due to disturbance. (For three-hourly range indices K see p. 213.)

11.6. The u-measure of magnetic activity. The inter-diurnal variability U of the horizontal intensity H for any station P is defined as the difference between the mean values of H for that day and for the preceding day, taken without regard to sign. Its average amount depends on the (magnetic) latitude or north-polar distance (Θ); an equatorial value u of the inter-diurnal variability is obtained by taking the mean of $U/\sin\Theta\cos\psi$ for a number of observatories, where ψ denotes the angle between the direction of H at P, and the great circle passing through P and the magnetic-axis poles; the division by $\sin\Theta\cos\psi$ is made in order to base u only on the component of H parallel to the earth's magnetic axis. This equatorial value u, expressed in units of 10γ, is taken as a measure of the magnetic activity. It is not appropriate as a measure for a single day, but its monthly or annual mean values measure the average amount of one aspect of magnetic disturbance during the corresponding periods. Clearly u is independent of the amount of change within each day, whether due to disturbance (as measured by R_D) or disturbance together with S_q (as measured by the absolute daily ranges); it is instead connected with the systematic effects of disturbance—shown chiefly in H (9.2)—both the 'direct' effects and the gradual recovery, or post-perturbation. Bartels [16–19] has calculated monthly and annual mean values of u for the period 1872–1937, using data from several observatories reduced to a common basis; he has also calculated annual means back to 1835, so that a 100 years' homogeneous series exists.

The lowest monthly value during the 756 months 1872–1934 is 0·38 (July 1913). Even in quiet times the fluctuations of H near the equator, in the monthly average, amount to more than 3γ change from day to day, which is a significant indication of the prevalence of minor disturbance. The highest monthly means are due to great magnetic

storms: e.g. for November 1882, $u = 2\cdot95$; January 1938, $u = 2\cdot74$; May 1921, $u = 2\cdot70$; March 1920, $u = 2\cdot54$; August 1917, $u = 2\cdot37$, coinciding with the highest monthly mean (154) of the relative sunspot number since 1872.
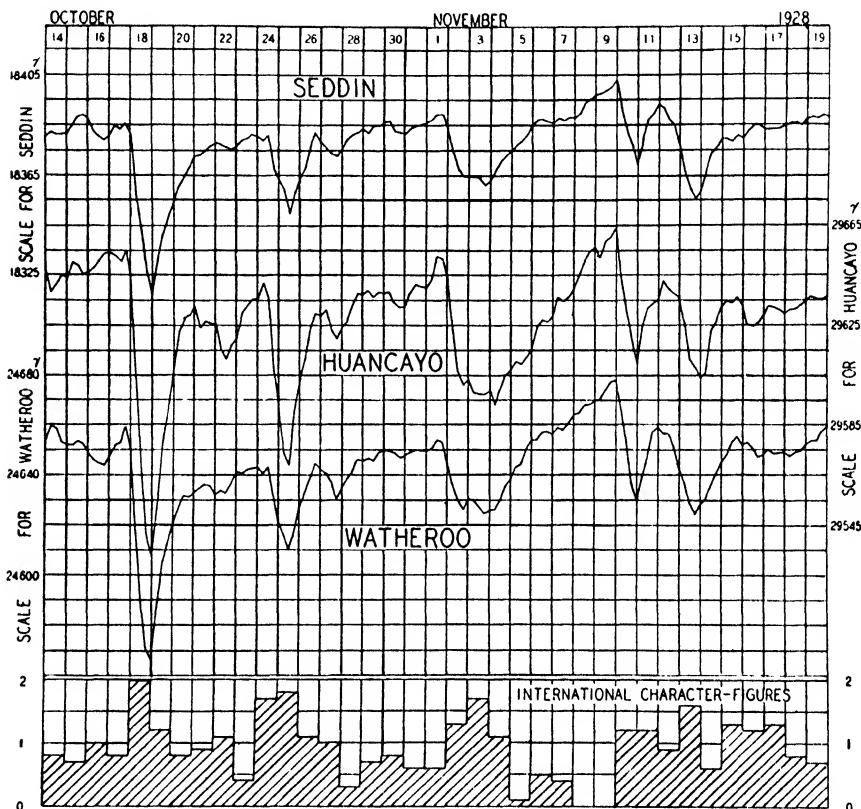
FIG. 5. Graphs showing the variation of 24-hourly mean values, centred at successive epochs 6 hours apart, of the horizontal magnetic intensity H (or north component X) at Seddin (Germany), Huancayo (Peru), and Watheroo (W. Australia). The vertical lines correspond to Greenwich midnight. The similarity of the three curves illustrates the basis of the u-measure

The frequency-distribution of positive and negative day-to-day changes (all to the nearest γ) of the north component X at Potsdam, 1905–20, expressed as the number per thousand, are given below; in addition, 79 changes per thousand were zero.

Change in γ .	1	2	3	4	5	6–8	9–12	13–20	21–40	>40	Sum
Positive . .	90	90	84	63	53	94	39	16	4·4	0·3	534
Negative . .	71	65	46	38	29	52	36	28	16·4	5·0	387

The asymmetry (although expected from the general morphology of disturbance) is striking; negative differences are less frequent than positive, but large decreases (negative changes), above 20γ, are nearly five times as frequent as large increases. The most frequent change is $+1\cdot5\gamma$. The highest individual changes during the 16 years' interval were $+74\gamma$ (to 1920 March 24) and -86γ (to 1908 September 12). W. van Bemmelen [20] has discussed the relation of the interdiurnal change of H, or X, to the character figures.

11.7. A statistical aspect of measures of activity. The frequency-distribution of the monthly means of u is characterized by many values near the lower limit (for 10 per cent. of all months, $u \leqslant 0\cdot52$), but the average value, $0\cdot85$, is greatly exceeded in some months with storms. This distribution, with occasional extremely high values, should be characteristic also for monthly measures defined as an average of daily ranges, and even more so if the ranges are squared before averaging. The monthly means of the international magnetic character figures C have, however, a different frequency-distribution, because the highest daily figures, $C = 1\cdot9$ or $2\cdot0$, do not allow further discrimination between storms exceeding a certain range. Consider, for instance, March 1920, for which $u = 2\cdot54$, and the influence of the storm of 1920 March 22 and 23 on the average interdiurnal variability U of the north component X at Seddin. The four days March 22 to 25 nearly double the average U, from $7\cdot8\gamma$ (for the remaining twenty-seven days) to $13\cdot6\gamma$; but they raise C only slightly, from $0\cdot65$ to $0\cdot79$.

11.8. The u_1-measure. For statistical purposes it appears desirable to obtain an objective measure of activity which shares with C the property of reducing the influence of the exceptionally great disturbances. Hence the modified u_1-measure was introduced, this being a function of the monthly mean of u, defined in the manner indicated by the following table of corresponding values, or rather by the continuous graph indicated by this table [16].

$u =$	0·3	0·5	0·7	0·9	1·2	1·5	1·8	2·1	2·7	3·6 and higher
$u_1 =$	0	20	40	57	79	96	108	118	132	140

This relationship was chosen so that the monthly means of u_1 should have a frequency-distribution similar to that of the sunspot numbers. Other relationships might, of course, have been chosen instead.

11.9. The annual variation of magnetic activity. The 59 years 1872–1930 were divided into three groups of 20, 19, and 20 years, of respectively high, medium, and low activity. The twelve *monthly* averages of u_1 were computed for each group, and their deviations from

the annual mean, expressed as percentages, are given in Table 1 and in Fig. 6. In addition, the annual variation of the character figure C for the years 1906–30 is shown in the same manner, in the last line. The double wave, with maxima near the equinoxes, is pronounced. The

TABLE 1. *Annual variation of magnetic activity u_1 and character figures C, expressed as percentage deviation from the annual mean values*

	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.	Mean
u_1, high	−13	+3	+12	+3	−7	−14	−6	+4	+8	+17	+4	−10	69
u_1, medium	−20	−7	+7	0	+5	−12	−14	0	+17	+9	+19	−3	51
u_1, low	−11	−4	+12	0	+9	−5	−16	−9	+12	+15	+7	−9	32
u_1, all	−15	−2	+10	+1	0	−12	−11	0	+12	+14	+10	−8	50
C, all	−3	+8	+13	−2	+1	−9	−8	−2	+10	+10	−9	−10	0·63
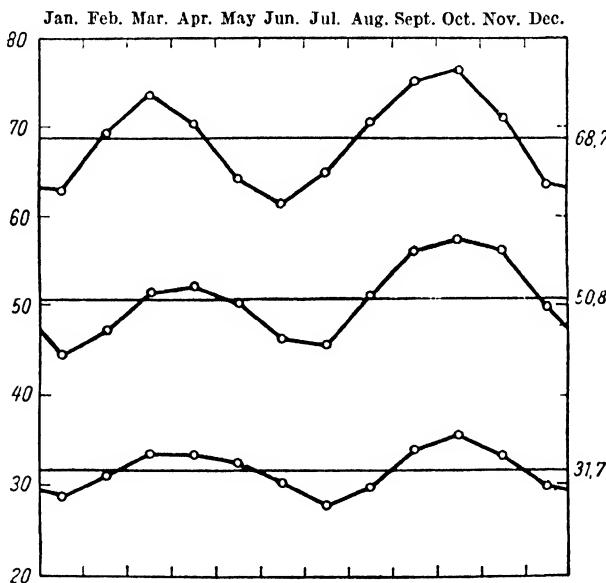
FIG. 6. The annual variation of the magnetic activity u_1 in years of high, medium, and low activity

average values of u_1 and C for the four equinoctial months March, April, September, October are 14 per cent. (in u_1) or 12 per cent. (in C) above the average for the remaining eight months; the corresponding excess for the disturbance-range R_D of §3 is 13 per cent. The annual variation in u_1 makes a considerable contribution (one-third) to the total fluctuation of u_1 (cf. 16.34). Table 2 gives the annual variation in the frequency of magnetic storms (p. 328), for comparison with Table 1.

Statistical considerations such as are described in Chapter XVI have.

TABLE 2. *Annual distribution of magnetic storms* 1875 *to* 1927, *listed in Greenwich Photoheliographic Results* 1927

	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.	Total
Great storms	3	6	8	5	6	4	4	10	7	3	3	1	60
Smaller storms	30	29	35	28	28	19	15	21	38	47	33	19	342

Average number per individual month:
months March, April, September, October: 0·108 great storms, 0·65 small storms
other months: 0·087   ,,   0·46   ,,

shown that the double wave is significant, but that there is no significant superposed twelve-monthly wave; in other words, the asymmetry in Table 1, which shows a higher activity near September than near March, can be explained as an accidental accumulation of irregular, non-seasonal fluctuations. No effect on u_1 of the varying distance between the earth and the sun (5.1) has been found.

The exact dates of the maximum of the six-monthly variation of disturbance are found by harmonic analysis to be April 5 and October 7 for u_1, March 16 and September 18 for C. These dates are of interest in connexion with attempts to explain the semi-annual variation of geomagnetic activity as depending on the sun's *geocentric* position (relative to the earth's equator), or alternatively on the earth's *heliocentric* position (relative to the sun's equator). The passages of the sun across the equator occur on March 22 and September 22 (the equinoxes); those of the earth across the sun's equator occur on March 5 and September 5. While statistical tests [16] favour the former dates, a satisfactory decision is not yet possible.

Some meteor showers recur at the same date every year. Corresponding annual recurrences of magnetic activity have been looked for, but the evidence on this point, provided by Maris [21–3], is not convincing [16].

11.10. Correlation. N pairs of associated values of two variables x_n and y_n ($n = 1, 2, ..., N$), with averages x_0, y_0, are characterized statistically by their standard deviations σ_x, σ_y and their correlation-coefficient r, such that

$$x_0 = \sum x_n/N; \qquad \sigma_x^2 = \sum (x_n - x_0)^2/N \equiv \sum x_n^2/N - x_0^2 \tag{5}$$

(and likewise for y_0 and σ_y), and

$$r = \sum (x_n - x_0)(y_n - y_0)/N\sigma_x \sigma_y \equiv [(\sum x_n y_n/N) - x_0 y_0]/\sigma_x \sigma_y. \tag{6}$$

The correlation-coefficient r is discussed in text-books on statistics [1–8]. For our purposes, its meaning is perhaps best illustrated by

considering the case where the relation between x and y is approximately linear, so that approximately $y_n = ax_n + b$. Suppose that a and b are determined so that the computed values

$$y_n' = ax_n + b \qquad (7)$$

differ from the observed values y_n as little as possible, this being interpreted as requiring that the sum of the squares of the residuals

$$\Delta y_n = y_n - y_n' \qquad (8)$$

shall be a minimum (for this reason the condition is called that of *least squares* (16.8)). The conditions for this minimum are easily seen to be that

$$a = r\sigma_y/\sigma_x, \qquad b = y_0 - ax_0. \qquad (9)$$

The latter equation implies that y_0 is the mean value of y' as well as of y. The standard deviation of y' from its mean value y_0 is $r\sigma_y$, and that of Δy_n from its mean value (zero) is $\sigma_y(1-r^2)^{\frac{1}{2}}$. In other words, y_n can be regarded as the sum of the two parts y_n' and Δy_n, of which y_n' is a strictly linear function of x_n, while the other part Δy_n is not correlated with x_n; the relative magnitudes of these parts, estimated by their standard deviations, stand in the ratio

$$r : (1-r^2)^{\frac{1}{2}}; \qquad (10)$$

for instance, this is $1:1$ if $r = 1/\sqrt{2} = 0 \cdot 707$.

If we interchange the roles of x and y, and regard x as depending approximately linearly on y, so that (7) is replaced by

$$x_n' = cy_n + d, \qquad (11)$$

the condition of least squares of Δx_n gives *not* $c = 1/a$, but $ac = r^2$, as can be seen from (9).

The discussion of the *irregularities* in the relation between x_n and y_n can conveniently be made by considering the corresponding *normalized* variables $(x_n - x_0)/\sigma_x$ and $(y_n - y_0)/\sigma_y$, which are expressed as deviations from the average values, and in such units that $\sigma_x = \sigma_y = 1$. For these variables $a = c = r$, and $b = 0$. If the normalized variables are conceived as coordinates (x to the right, y upward) of a cloud of points in a plane, the straight line corresponding to (7) and (9) has the slope r relative to the x-axis, and the sum of the squares of the *vertical* distances of the points from the line is a minimum; the (different) straight line corresponding to (11) has the slope r relative to the y-axis, and the sum of the squares of the *horizontal* distances is a minimum. (For further general discussion see § 12.)
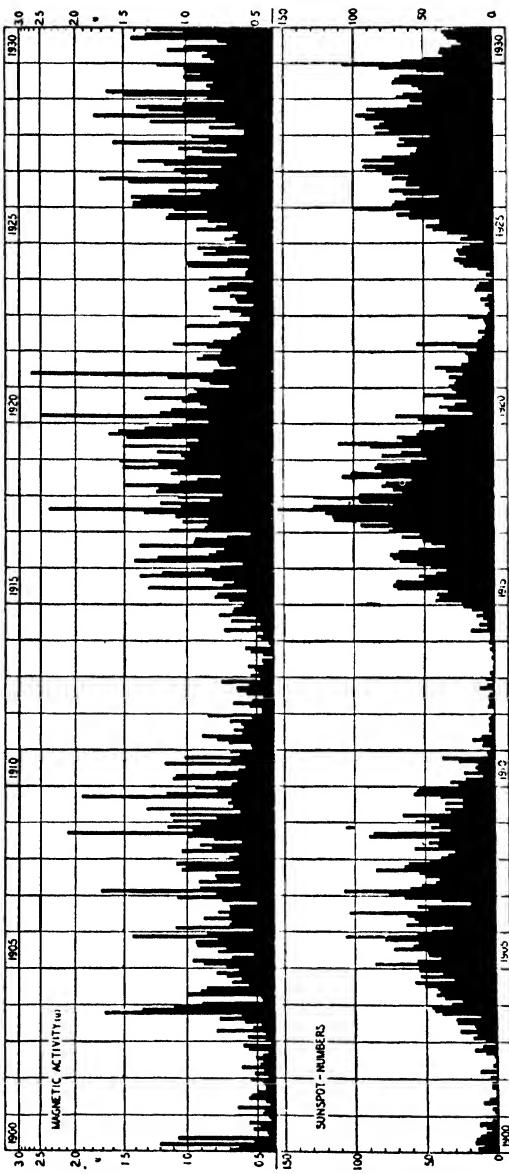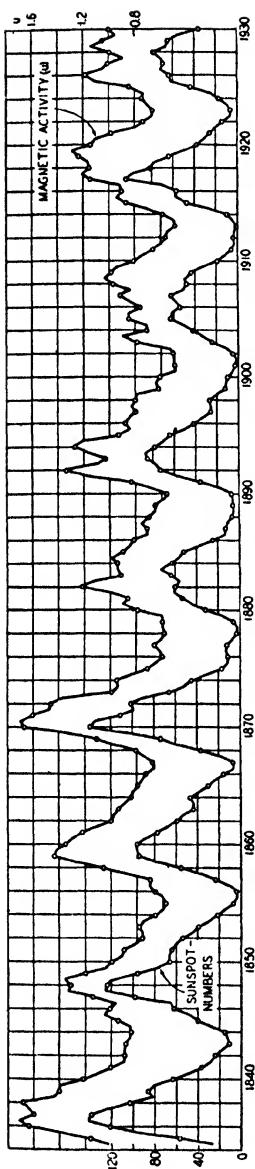
Fig. 7. The magnetic activity u and the sunspot number; annual means, 1834–1930, above, anu monthly means, 1900–30, below

11.11. Correlations between annual means of magnetic and solar activity. These considerations will now be applied to the annual mean values of the relative sunspot numbers s and of the magnetic activity measures u and u_1, 1872–1930. Their respective average values are 38·9, 0·854, and 50·4; their standard deviations are 27·5, 0·221, and 16·9. The correlation-coefficients between s and u or u_1 are very high, namely,

$$r = +0\cdot869 \text{ between } s \text{ and } u, \qquad r = +0\cdot884 \text{ between } s \text{ and } u_1.$$
$$(12)$$

The ratios (10) are 1·76 and 1·89 respectively. As might be expected,
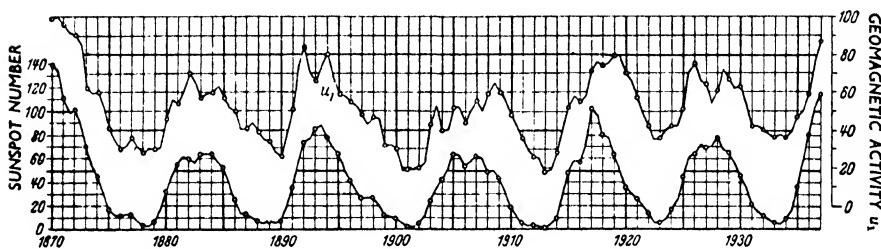
Fig. 8. The magnetic activity u_1 (upper curve) and the sunspot number (lower curve), annual means, 1870–1937

the correlation between u_1 and s is slightly better than that between u and s (see Figs. 7 and 8).

Similar calculations have been made also for the more limited period 1906–30, for which values of the magnetic character figures C were available. For the annual mean values, as before, it is found that

$$r = +0\cdot570 \text{ between } s \text{ and } C, \qquad +0\cdot820 \text{ between } s \text{ and } u_1,$$
$$+0\cdot719 \text{ between } C \text{ and } u_1. \qquad (13)$$

This indicates the inferiority of the *annual* means of C to those of u_1, as a measure of the yearly average of terrestrial-magnetic activity, because these correlation-coefficients show that u_1 is better correlated with the sunspot numbers s than C is, or even than u_1 is correlated with C, that is, with another but less satisfactory measure of the magnetic activity itself.

The annual numbers of sunspot areas, faculae, and limb-prominences do not give significantly higher correlations with u_1 than the sunspot numbers s; this was to be expected, because all the measures of solar activity are highly correlated with one another; for instance, the annual values of the sunspot areas and relative sunspot numbers, over the period 1882–1930, yield $r = 0\cdot981$, with the ratio (10) as high as 5·31.

Brunner [24] has published a series of monthly means of the profile areas of the solar limb-prominences from 1910 to 1932; ninety-two quarterly means (for three months each) were formed, and this measure (say P) was compared with the quarterly sunspot numbers s and the quarterly means of the magnetic activity u_1. The correlation-coefficients are: $r(u_1, P) = 0·711$, $r(u_1, s) = 0·739$, $r(P, s) = 0·840$.

The approximate linear relation (1872–1930) between u_1 and s is

$$u_1 \doteq 0·543s + 29·3 \equiv u_1'. \tag{14}$$
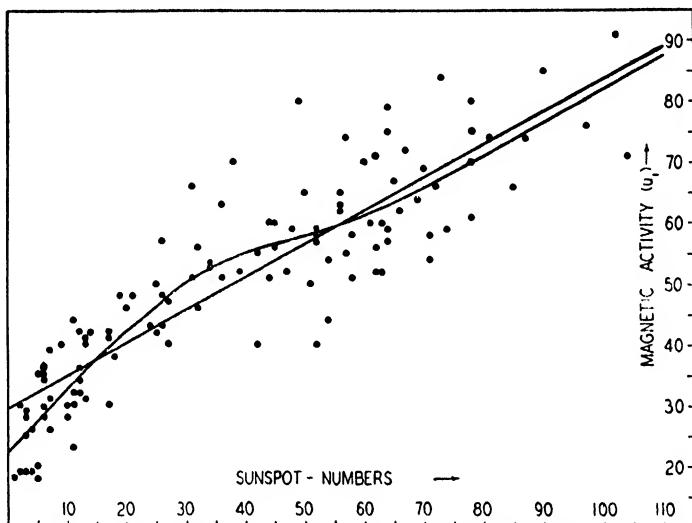
FIG. 9. The magnetic activity u_1 and the sunspot number; annual means

Closer inspection of the residuals $\Delta u_1 = u_1 - u_1'$ shows that u_1 responds to changes of s more strongly when s is small than when s is large; for instance, u_1 increases by 20 units from $s = 0$ to $s = 20$, instead of by 11 units as (14) indicates. This also appears from Fig. 9.

The sensitivity of u_1 for small changes of s near sunspot-minimum was noticeable in 1923 ($u_1 = 36$), when u_1 did not fall to the low values of the two preceding sunspot-minima, 1901 and 1913 ($u_1 = 19$ and 18), because '1923 did not have so pronounced a character, as a sunspot-minimum year, as 1901 and 1913' (Wolfer).

A certain lag of u or u_1 behind s in the 11-year cycle is apparent since 1900 (Figs. 8 and 9), Δu_1 being negative in the ascending and positive in the descending phase. This can also be expressed by the statement that, in the *descending* phase of the cycle, the same value u_1 is connected with values of s that are 20 units lower than those

corresponding to the same value u_1 in the *ascending* phase. A similar phenomenon was discussed in § 3. It supports the idea that magnetic disturbance is due to a more or less radial emission from the sun, which is more likely to sweep across the earth during the descending phase of the cycle, when sunspots are nearer the sun's equator, than when their solar latitude is higher; but, of course, it may be due to some other change on the sun, which also progresses with the sunspot cycle.

In Figs. 7 and 8, the sunspot-minimum years 1901 and 1913 appear appreciably quieter, magnetically, than the minima of 1923 and 1933. This magnetic evidence agrees with solar observations in so far as the sun was free from spots on more days in 1901 and 1913 than in 1923 and 1933. The percentages of days on which sunspots were observed, in the three years centred at each sunspot-minimum, were as follows:

Years	1900-2	1912–14	1922–4	1932–4
Days per hundred	34	36	66	57

These figures characterize the difference in solar activity, and accord better with the geomagnetic activity than the respective annual sunspot numbers for the minimum years, which are as follows:

Minimum year	1901	1913	1923	1933
Annual sunspot number	2·6	0·0	5·6	3·4

11.12. Correlations between monthly means. The monthly means of u_1 and s are decidedly less correlated with one another than are the annual means. For the period 1872–1930 the correlation coefficient for the monthly means of u_1 and s is

$$r = +0.654, \tag{15}$$

as compared with $+0.884$ for the annual means. For the deviations of the monthly means from their respective annual means, the correlation is only $+0.292$, with a ratio (10) equal to 0.30 only. This indicates that u_1 follows s closely in the main trend of the 11-year cycle, but that the fluctuations in s from month to month are little reflected in u_1.

This is a typical example of a correlation in a series characterized by conservation-effects (16.27). Consider N pairs of annual means A_n and B_n (with averages A_0 and B_0), derived from NM monthly means a_{nm} and b_{nm} ($M = 12$), or NM monthly departures α_{nm} ($= a_{nm} - A_n$) and β_{nm} ($= b_{nm} - B_n$). Let the standard deviations be denoted by σ_A, σ_B; σ_a, σ_b; σ_α, σ_β. Then it can be shown by simple application of formulae of the type

$$\sum A_n B_n = \sum (A_n - A_0)(B_n - B_0) + NA_0 B_0 = N(\sigma_A \sigma_B r_{AB} + A_0 B_0), \tag{16}$$

that
$$\sigma_a^2 = \sigma_A^2 + \sigma_\alpha^2, \qquad \sigma_b^2 = \sigma_B^2 + \sigma_\beta^2, \tag{17}$$

and
$$\sigma_a \sigma_b r_{ab} = \sigma_A \sigma_B r_{AB} + \sigma_\alpha \sigma_\beta r_{\alpha\beta}. \tag{18}$$

If
$$p_\alpha \equiv \sigma_\alpha/\sigma_A, \qquad p_\beta \equiv \sigma_\beta/\sigma_B, \tag{19}$$
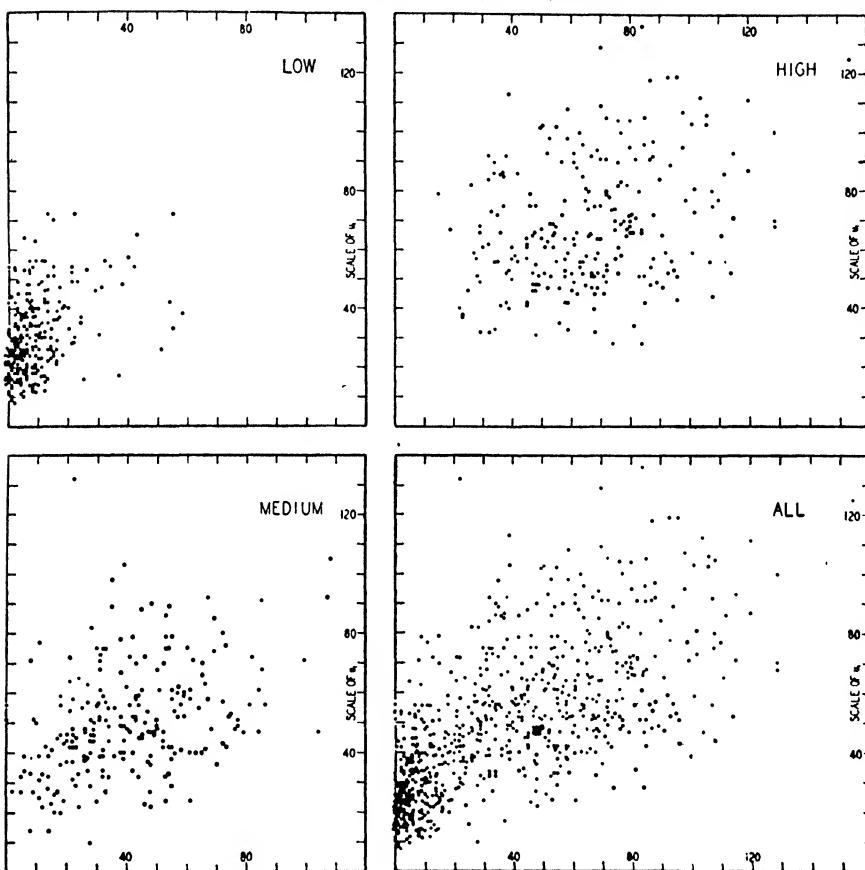
FIG. 10. The relation between monthly mean values of the magnetic activity u_1 (increasing upward) and the relative sunspot numbers (increasing to the right), (i) for 20 years of *low* activity, $u_1 \leqslant 42$, (ii) for 20 years of *high* activity, $u_1 \geqslant 58$, (iii) for 19 years of *medium* activity u_1, and (iv) for *all* the 59 years, or 708 months

(18) may be written
$$r_{ab} = (r_{AB} + p_\alpha p_\beta r_{\alpha\beta})/\{(1+p_\alpha^2)(1+p_\beta^2)\}^{\frac{1}{2}}. \tag{20}$$

This can be most easily illustrated by a diagram representing the data by clouds of points, having the rectangular coordinates A_n and B_n, or a_{nm} and b_{nm} (Figs. 10, 11). Each point A_n, B_n is the mass-centre of M points a_{nm}, b_{nm}, or, vice versa, the point-cloud for the monthly means is obtained by an *explosion* of the points for the annual means

[16]. It is the magnitude and direction of this 'explosion' which determines p_α, p_β, $r_{\alpha\beta}$, and consequently also the reduction of the annual mean correlation r_{AB} to that of the monthly mean correlation, r_{ab}, as in (20). (See also 19.2.)

In conclusion, it must be mentioned that in certain selected groups of years, at the end of a sunspot-cycle, the correlation between the monthly means of s and u_1 is especially poor [16]. For instance,
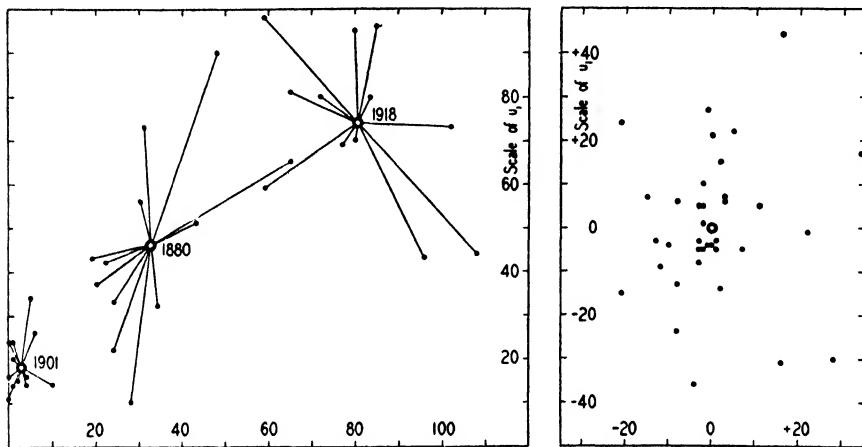
FIG. 11. On the left: annual and monthly values of the magnetic activity u_1 (increasing upward) and the relative sunspot number (increasing to the right), for three separate years of low, medium, and high activity. On the right: the corresponding monthly departures from the annual means. To illustrate the 'explosion' relation between the correlation-coefficients for the annual means, monthly means, and monthly departures

for the thirty-six monthly means 1918–20, $r = +0.06$, and for 1928–30, $r = +0.01$. This remarkable obliteration of the otherwise close relations between sunspots and magnetic activity in certain intervals is not removed by the use of other measures of solar activity, and therefore suggests caution in drawing conclusions from observations over a limited number of years. The amount of independent material available appears highly reduced if the 'conservation' effect is taken into account (16.27 et seq.).

11.13. Disturbance statistics at individual stations. It has been seen in 6.3 that disturbance is a world-wide and not merely a local phenomenon. It is manifested in all magnetic elements, though in different amounts, the ratios of which vary with the locality. It is therefore to be expected that many of the results obtained for the disturbance-range R_D in Greenwich declination will be paralleled in other elements and at other stations.

But even if a given set of days ranged themselves in exactly the same order, according to their disturbance as judged by R_D in two different elements and/or at two different stations—which is unlikely— the changes in R_D would not necessarily be in a *constant* ratio. The ratio might vary with the general intensity of disturbance, with the season, with the sunspottedness, and so on. If so, though the two criteria might lead to frequency curves, constructed in the same way as those of Figs. 1, 2, 3 for Greenwich declination, which resembled the latter in general form, the corresponding curves would not be geometrically similar, and the differences between the curves for selected groups of days might not be similar in the two cases. Likewise the variations in the *mean* values of R_D for such groups of days might be different. These various possibilities can be tested only by detailed investigation.

This has been done for various stations and elements in middle and low latitudes, and it appears that all of them show an annual variation, and a variation of the annual means throughout the sunspot cycle, of the same general type as that obtained for the Greenwich declination. But the amounts of these variations relative to the mean value of R_D in each case, and also these means themselves, vary with the latitude.

The annual variation of magnetic disturbance, whether measured by R_D or by u, has been considered for observatories in different latitudes. The annual variation is not the same in low and high latitudes, indicating that the distribution of disturbance over the earth depends on the geometrical relationship of the earth to the sun; the sunspot numbers, which represent intrinsic solar changes, naturally show no annual variation [16].

The results indicate that in all its components the disturbance variation-field D, throughout middle and low latitudes, has a seasonal change in which the maxima of intensity occur at the equinoxes, and the minima at the solstices: and that it undergoes a variation in the course of the sunspot cycle, in which its intensity follows the average changes of sunspottedness.

11.14. Disturbance statistics at polar stations. For stations in polar latitudes the available magnetic record is usually insufficient to indicate whether there also the field D varies in intensity with the mean sunspottedness. But the close association of magnetic disturbance (everywhere) with auroral phenomena (Ch. XIV), and the known correlation between the latter and the annual mean sunspottedness, makes

it probable that the intensity of the D field in polar as well as in lower latitudes follows the sunspot variations.

The case appears different with regard to the seasonal change of disturbance-intensity. Chree [25, 6.5, 9.3] has shown that at Cape Evans (77·6° S., 166·4° E.) in the Antarctic, the summer (there centred at the December solstice) is the most highly disturbed season. His material consisted of the *total* daily ranges R in declination and vertical force. These ranges are affected by the quiet-day solar daily variation S_q, as well as by disturbance D, and their frequency-distribution therefore depends on two distinct phenomena. But S_q (on really quiet days) is relatively small in the polar regions, so that R and R_D will be nearly equal. The following are the percentage frequencies of days within which R lay within certain ranges:

TABLE 3. *Antarctic. Declination daily ranges*

Range	0′–30′	30′–60′	60′–120′	120′–180′	Over 180′
Midwinter . .	15	37	25	15	8
Equinox . .	2	18	49	19	12
Midsummer . .	0	1	18	39	42

Vertical force daily ranges

Range in γ	0–50	50–100	100–200	Over 200
Midwinter . . .	38	43	16	3
Equinox . . .	20	48	29	3
Midsummer . . .	6	21	52	21

These results indicate that the geographical distribution of intensity of the D field changes with the season. The intensity is probably always greater in the polar regions than in lower latitudes, but at the June solstice the northern polar (geographical) maximum of intensity is greater than the southern polar maximum, while at the December solstice the order is reversed. At the equinoxes the two polar maxima are probably of equal magnitude, and at this season the intensity in middle and low latitudes is greater than at the solstices.

11.15. The lack of general correlation between the sunspot numbers and the magnetic activity on individual days. The preceding discussion has shown that there is a close relationship between the yearly average values of sunspot numbers and of magnetic disturbance, whether the latter is measured by R_D or by u or u_1. The relationship between the *daily* sunspot numbers and the values of the magnetic activity on the same day is, however, by no means so close.

This is suggested at once by the curve for the sunspot group of years ϵ, in Fig. 2; the annual mean sunspot number for the years in this group is only 6, and on many of the days in these years the sunspot number must have been zero (as indicated by Fig. 5.8, p. 175); yet many days in the group show considerable disturbance-ranges R_D, and the mean value of R_D for the group (4·18′) is more than half that (7·16′) for the sunspot-maximum group α.

The weakness of the relationship between sunspottedness and disturbance on the same day has been demonstrated by Chree [28–9] in many striking and varied ways. In the first place, he showed that the average sunspot number (41·0) for all days during the 11-year period 1890–1900 was the same as the average for the five quiet days per month (660 in all, during these years) as selected by the Astronomer Royal (which gave 41·1 as average).

The equality of the mean sunspot numbers for the two groups of days is in striking contrast with the difference between the corresponding mean absolute daily ranges (which include S_q) in Kew declination, which were 9·6′ for the quiet days and 13·6′ for all days.

Another illustration given by Chree referred to these quiet days (or rather to 565 of them), and also to 191 out of 209 days (during the same period) which had been selected at Kew as highly disturbed. All the days of each month of the eleven years were divided into three classes, A, B, and C, according to the corresponding sunspot areas; A contained the 10 days of largest spot area, C the 10 days of least area, and B the intermediate days. Nineteen months during the quiet years 1890, 1899, and 1900 were omitted as containing each more than 10 days free from spots. The following table shows how the 565 quiet days and 191 disturbed days (omitting those which fell during the nineteen months) were distributed between the sunspot groups of days A, B, C.

Number of quiet days			Number of disturbed days		
A	B	C	A	B	C
179	195	191	68	65	58

If magnetic disturbance and sunspot area on the same day had been absolutely unconnected, it would be expected that 188 quiet days and 64 disturbed days would have fallen in each group A, B, C; the table shows that this is nearly but not quite the case, there being a deficiency of quiet days in the group A of largest spot area, and an excess of disturbed days in the same group; but both the deficiency and excess

are very slight, although the mean sunspot number for group A was 5·2 times that for group C.

Chree also investigated the excess of the daily range in Kew declination over the mean for the whole period 1890–1900, for the days of group A and the first, second, and third following days, and likewise for the days of group C and the three following days. The results were as follows:

	Group days	First days after	Second days after	Third days after
Sunspot group A .	+0·17′	+0·25′	+0·48′	+0·53′
„ „ C .	−0·32′	−0·45′	−0·38′	−0·35′

These numbers suggest that the magnetic disturbance conditions are related very slightly, and almost equally, to the solar spottedness on the same day and on each of the three previous days; large spot areas are associated with slightly larger, and small spot areas with slightly smaller, magnetic ranges. (In individual years the average range-differences were larger than for the mean of the eleven years, and sometimes were opposite in sign to the above mean values.)

In addition to further similar illustrations, Chree pointed out a number of special cases illustrating the absence of a necessary and invariable connexion between magnetic disturbance and sunspottedness; more striking examples will, however, be given in § 17.

Further evidence for the apparent lack of a statistical relationship between magnetic and solar activity on individual days will be given in Chapter XII.

The inconclusive results discussed in this section are due to the prevalence of minor disturbance. Intense magnetic storms, on the contrary, are decidedly correlated with individual large sunspots, and vice versa.

11.16. Large magnetic storms and large sunspots. On the basis of the Greenwich observations, Greaves and Newton [30, 31], continuing similar work by Ellis [32, 33] and Maunder [34], prepared two independent catalogues for the interval 1874–1927, one including all the 'naked-eye' spot groups (of mean area at least equal to 500 millionths of the sun's hemisphere), and one including all the magnetic storms with ranges greater than 0·5° in declination, or 150 γ in H and Z at Greenwich. The first catalogue numbered 455 spots, the second 403 storms. The storms were also classified into three sets, of 343 smaller storms, 43 great storms, and 17 very great storms, the dividing criteria

being 1·0° and 1·5° in declination, and 300 γ and 500 γ for the intensity components; the 23 largest spots, with mean areas of over 1,500 millionths, were also noted. A storm and a naked-eye spot were called 'associated' if the storm began within the eight-day interval centred at the time of the passage of the spot across the central meridian of the sun. (Some of these eight-day intervals overlapped with another, but even if all had been quite separate, they would have included less than one-fifth of the whole time-interval (fifty-four years) considered.) It was found [31 a] that:

   16 of the 23 largest spots were associated with one of the 403 storms;

   21 of the 43 great storms were associated with one of the 455 spots;

   15 of the 17 very great storms were associated with one of the 455 spots.

These figures are certainly greater than could be expected in a chance distribution.

In the interval 1875–1927, five minima of the sunspot cycle occurred. In the five periods each of three years, surrounding each minimum, there occurred only 3 great or very great storms out of the 60, and 53 smaller storms out of the 343; a uniform distribution over the fifty-three years should have given (15/53) times 60 or 343, that is, 17 or 97 respectively. The storms are therefore relatively rare in the years near sunspot-minima.

The different degrees of magnetic activity near the individual sunspot-minima, mentioned in § 11, also appear in the total number of storms in the three-yearly groups of minimum years, namely, 5 in the years 1900–2, 4 in the years 1912–14, but 14 in the years 1922–4.

For each storm, the spots which were within four days (53° of heliographic longitude) on either side of the central meridian of the sun at the time of the commencement of the storm were examined; the spot with the largest area was selected (if there was more than one) and its mean area was written down against the storm. The largest spots of those near the limb (between four and six days east or west from the central meridian) were similarly selected. For comparison, the corresponding mean areas for 264 very quiet days are given, obtained by selecting, in each month of the years 1914–24 inclusive, the two days with the lowest international magnetic character figure. The mean

areas (all corrected for perspective foreshortening) in each group are as follows [31 b]:

	Mean areas of the largest spots near the centre, 53° E. to 53° W.	Mean areas of the largest spots near the limb, 53° to 79° E. and W.
264 quiet days	244	167
343 smaller storms	324	203
43 great storms	691	251
17 very great storms	1,116	181

The two figures in each row are not directly comparable, because the centre group relates to 106° heliographic longitude, and the limb group only to 52°; but the *increase* in the first column, from the quiet days to the days of greater storminess, is so much more significant than the variations in the last column as to leave no doubt that storms are often in some way associated with the presence of spots near the central meridian.

A central sector between 79° east and west was further subdivided into six sectors of 26° longitude, and the mean areas of the largest sunspot in each of these sectors of heliographic longitude were computed for the quiet days and for the three groups of storms; in the latter case the sectors refer to the epoch of commencement of the storm. The results were as follows. (Note that the first three columns relate to sectors which have not yet crossed the central meridian.)

TABLE 4

	E. 79°–53°	53°–27°	26°–0°	0°–26°	27°–53°	53°–79° W.
Quiet days	85	79	82	89	91	109
Smaller storms	91	76	131	120	109	136
Great storms	115	119	203	306	308	176
Very great storms	122	172	460	380	445	78

The largest increase appears in the third, fourth, and fifth columns. This suggests that the occurrence of large magnetic storms is associated with the presence of large spots in a sector between 26° E. and 53° W. The average position is a meridian which, at the time of the commencement of the storm, had passed the central meridian about one day before.

There are, however, some notable exceptions to the association of large storms and large sunspots:

(a) The very great storm of 1894 November 13–14, with a range of 500γ in vertical force, occurred at a time when there was no unusual activity visible on the sun's disk, and indeed when the solar activity was far below the average for the year.

(b) Thirty-two of the smaller storms commenced on spotless days.

(c) The storm of 1901 May 10 had a sudden commencement; no spots or faculae were measured at Greenwich between 1901 March 28 and May 18.

(d) A large sunspot, of 1,838 millionths mean area, crossed the sun's central meridian on 1905 July 17; no appreciable magnetic disturbance occurred for several days before or afterwards.

Summarizing, Greaves and Newton find that the solar disturbances responsible for magnetic storms, and probably situated in the chromosphere, possess the following characteristics:

(1) their frequency changes with the 11-year cycle;

(2) the larger of such disturbances are likely to be accompanied by sunspots.

The available data seem to indicate (although not decisively) that storms with sudden commencements are more closely correlated with sunspots than storms without this characteristic.

11.17. Spectrohelioscopic observations, individual cases. The fact that there are exceptions to the general connexion between magnetic storms and the simultaneous presence, size, and position of sunspots led Tacchini and Hale to suggest that the storms may be due to the eruptive phenomena often found (especially with the aid of the spectroheliograph) near spots. Unfortunately, until recently it has been difficult to follow the development of these eruptions regularly and continuously, and doubtless many are still overlooked: a few coincidences between them and the recurrence of magnetic storms have, however, been noted. An intense solar eruption which was practically simultaneous with the outbreak of a short fluctuation, and which was followed, about 18 hours later, by a considerable terrestrial magnetic disturbance, was observed on 1859 September 1 by Carrington and Hodgson (9.23); and Young noted 'peculiar twitches of the magnets in England' on 1872 August 3 and 5, simultaneous with brilliant distortions of the hydrogen line near the limb. On 1892 July 5 Hale photographed an exceptional eruption which varied rapidly both in shape and intensity; this was followed 21·3 hours later by the outbreak of a magnetic storm, which attained its maximum intensity in 4 hours. On 1908 September 10 Fox and Abetti observed a remarkable series of eruptive changes occurring near two sunspots, adjacent to one another but in different (northern and southern) hemispheres; the principal phase occurred at $7\frac{1}{2}^h$ G.M.T., when the photographs showed a violent outburst of hydrogen; after about 3 hours the eruption had subsided.

A great magnetic storm began on the earth next day at $9^h 47^m$ G.M.T., that is, 26 hours after the hydrogen eruption; the storm attained its maximum intensity in 17 hours. Other similar coincidences were observed on 1909 September 24 and on 1926 October 15, the intervals between the solar eruption and the commencement of the storm being 26 and 31 hours.

A summary of such individual cases, by G. E. Hale [5.3], has been extended by Maurain [37, 38]. The recent intensification of the watch on the sun by means of the spectrohelioscope has resulted in more cases being noted, of apparently related solar eruptions and magnetic disturbances; the violent magnetic disturbance of 1938 April 16, for instance, was preceded, at an interval of about 21 hours, by a solar eruption of unusual brightness (9.22). (See also p. 395.)

11.18. Statistical comparisons of sunspots and magnetic activity by the 'superposed-epoch method'. The connexion between sunspots and magnetic disturbance has been further investigated by a method which may be referred to as the *'superposed-epoch' method*, devised by Chree (cf. Ch. XII) for the study of the 27-day recurrence tendency. In a notable application of the method to the investigation of the relation between sunspots and magnetic activity, Maurain [37] used magnetic data for the forty-one years 1883–1923, from the Paris observatory, at Parc St. Maur till 1900, and subsequently at Val Joyeux. During this period a daily character figure, varying from 1 to 7, was assigned to each day; the figures 1 and 2 referred to quiet days, and the figures 5, 6, and 7 to days of notable disturbance: there were 855 days during the 41 years, to which the figures 5 to 7 were applied, or about 21 per year. These days were therefore much more disturbed than the 60 per year used in the studies of the 27-day recurrence tendency by Chree (whose choice of 5 most disturbed days in each month must have much reduced the standard in quiet months), though they were less disturbed than the storms (about 13 per year) considered by Maunder, Greaves, and Newton. The mean sunspot number s_0 was computed for these selected days, and also the means (s_n) for the 5 preceding and 3 following days ($n = -5$ to $n = 3$); these means were computed for each year, and for the whole period, and also for 10 years of low solar activity (1888–90, 1900–2, 1912–14, 1923). The ratios s_n/s_0 are given in the following table, which also includes results for the two periods considered by Chree (Ch. XII), and for two groups of years, six (A) of ascending sunspottedness (1891, 2; 1903, 4; 1915, 6) and twelve (D) of descending sunspottedness (1885, 6; 1894–8; 1908, 9; 1919–21).

TABLE 5. *Maurain's table of sunspot numbers for days of magnetic disturbance and adjacent days*

	Annual mean s	Mean s_0	Average no. of selected days per year	s_n/s_0								
				$n =$								
				-5	-4	-3	-2	-1	0	1	2	3
41 years .	36·9	40·1	21	1·038	1·064	1·077	1·074	1·041	1	0·984	0·974	0·963
10 quiet years .	5·9	6·3	12	1·120	1·346	1·292	1·180	1·084	1	0·975	0·941	0·848
1890–1900 .	41·8	45·2	26	0·945	1·000	1·022	1·026	1·003	1	0·966	0·952	0·954
1900–10 .	45·4	55·2	17	1·124	1·098	1·106	1·147	1·052	1	0·959	0·933	0·889
6 A years	46·5	47·1	25	1·010	1·060	1·084	1·095	1·036	1	1·029	1·026	1·013
12 D years	44·5	49·3	26	1·008	1·039	1·054	1·059	1·029	1	0·969	0·965	0·975

These numbers seem to leave no room for doubt that, on the average, on days of notable magnetic disturbance the sunspottedness is declining from a maximum value attained a few days before. The interval in days between the selected day and the preceding maximum of sunspottedness is about $2\frac{1}{2}$, $3\frac{3}{4}$, $2\frac{1}{2}$, $2\frac{1}{4}$, $2\frac{1}{2}$, $2\frac{1}{2}$ for the six sets of years chosen; it appears to be longer in years of low solar activity. In each set of years the sunspottedness on the selected days is above the average, but only slightly so. The percentage increase of s_n over s_0, on the day of maximum sunspottedness (preceding the selected day by 2, 3, or 4 days) appears to vary from year to year, but to be markedly increased near sunspot-minimum; the percentages for the six sets are about 8, 36, 3, 15, 10, and 6; the actual increases are 3, 2, 1, 8, 5, and 3, to the nearest unit. These results are the more remarkable when it is considered that sunspots are at least not the sole cause of magnetic disturbance, as is indicated, among many other ways, by the occurrence of disturbance, especially in quiet years, during periods free from sunspots for a period of 5 days before and 3 days after (Maurain found 30 such cases among his 855 disturbed days).

Stagg [39] considered the mean day-to-day change of the sunspot projected area A (reckoned in millionths of the sun's visible disk) for the intervals extending from 7 days before to 1 day after 78 selected disturbed days, during the years 1901–10; they were days of range exceeding 30′ in Kew declination. He found the following values of $A_n - A_0$, where A_n denotes the mean sunspot area for the nth day, and A_0 the mean for the selected days. These figures show that the maximum sunspot area occurred about $1\frac{2}{3}$ days before the selected days.

Day: $n =$	-7	-6	-5	-4	-3	-2	-1	0	1
$A_n - A_0$	-356	-275	-188	-79	16	71	64	0	-25

He also computed similar figures, as follows, for 600 disturbed days (the 5 most disturbed days per month) in the same period.

Day: $n =$	-7	-6	-5	-4	-3	-2	-1	0	1
$A_n - A_0$	-55	-32	-6	8	15	23	15	0	-28

The maximum area here occurs 2 or $2\frac{1}{4}$ days before the selected days: the excess over that on the selected days is much less than for the 78 more highly disturbed days, as was to be expected.

YEARS WITH MANY SUNSPOTS

YEARS WITH FEW SUNSPOTS
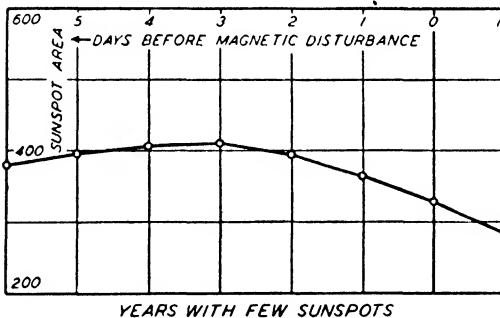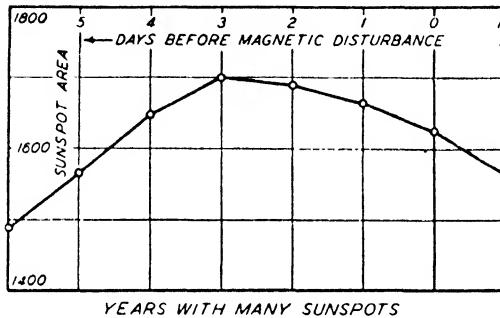
FIG. 12. Average sunspot areas, expressed in millionths of the sun's disk, from 6 days before until 1 day after days (epoch 0) selected as magnetically disturbed. Above: averages for 250 sets of days in years with many sunspots. Below: averages for 116 sets of days in years with few sunspots. (After J. M. Stagg)

Fig. 12 shows that in years with many sunspots the maximum of $A_n - A_0$ is more pronounced than in years with few sunspots.

11.19. The daily variation of magnetic activity. So far the discussion in this chapter has related to the occurrence or amount of disturbance throughout intervals of a year, a month, or a day. We have shown how the general intensity of disturbance varies with solar

sunspottedness and with the season, and how at each season it is dis-
tributed with respect to latitude. It is of interest also to consider how
it is distributed with respect to longitude relative to the sun, that is,
with respect to local time. For this purpose it is necessary to consider
the degree of disturbance during periods *less* than a day, e.g. for hourly
intervals.

In middle and low latitudes the mean intensity of disturbance aver-
aged over a day or a number of days is on the average similar for
stations at different longitudes λ in the same magnetic latitude. But
at any one station there is a *daily* variation in the intensity of *irregular*
disturbance (or of magnetic activity a, say), and for stations in the
same (middle or low) magnetic latitude the daily variations of a are
approximately the same at corresponding local times. In this respect
the daily variation of a is similar to the daily variation of the average
intensity of the D field (Ch. IX), and both of them are most con-
veniently thought of as representing one aspect of the geographical
distribution of a or ΔF relative to the noon meridian. The distinction
between a and ΔF should, however, be held clearly in mind; in con-
sidering a we refer to the intensity or relative frequency of the *irregular*
disturbance which we supposed averaged out when, in Chapter IX, we
considered the morphology of the average D field.

To investigate the change of disturbance-intensity throughout the
day, Chree [6.5] assigned a numerical character figure 0, 1, or 2 to each
hour of the day for certain years' observations at Eskdalemuir (Scot-
land) and at the Antarctic station of the second (British) Scott expedi-
tion. The figures were assigned having regard to the oscillatory
character of the traces for the three magnetic elements during the hour,
mental allowance being made for the regular daily variation. Just as
in the case of the international daily character figures, the standards
applied in different months and years were not the same, the primary
object being to discriminate between the various hours during each
individual month. The average character figure for each hour of the
day in seasonal groups of days was then determined, giving a daily
variation of character figure (treated as a measure of disturbance). The
daily variation of disturbance-intensity was also illustrated by con-
sidering the percentage of 2's, or 0's, among the character figures for
each hour.

The daily variation of irregular disturbance a was brought out fairly
clearly and consistently by the two methods, and indicated that at
Eskdalemuir the most disturbed period is from about 16^h to 20^h local

time, and the least-disturbed period from 6^h to 10^h, while at the Antarctic station the converse is approximately the case. Chree pointed out that as the two stations are on nearly opposite meridians, this means that their daily variations are similar relative to Greenwich time; but as is shown by Stagg's work on the daily variation of a at Arctic stations (*vide infra*), this is a function of the local time in high as well as in low latitudes.

To study the daily variation of a at Kew, Stagg [41] assigned hourly character figures 0, 1, and 2 for every hour throughout the eleven years 1913–23, by consideration of the Kew D and H magnetograms: the figures were assigned so as to characterize the divergence from the mean quiet-day variation for the season of the year concerned. Owing to the smaller interval characterized, a more precise scale of disturbance was possible than in the case of the daily character figures 0, 1, or 2 assigned at Kew; but the standard for a 2 (and presumably therefore for a 1 also) was consciously changed from sunspot-maximum to sunspot-minimum years.

The variation of mean hourly character figures throughout the day, for the three seasons and the whole year, is shown in Fig. 13. The daily variation is well marked, with a maximum at 0–1^h at all seasons: the hour of minimum varies with the season, being at about 10^h in winter, 9^h at the equinoxes, and 8^h or 9^h in summer, the annual mean minimum being at about 9^h. In summer there is a marked secondary maximum in the afternoon, at about 16^h. The range of the daily variation of disturbance, as thus measured, is greatest at the equinoxes; in winter it is slightly less than in summer. The mean hourly character figure is greater at the equinoxes than at the solstices for every hour of the day except from 13^h to 18^h, when, owing to the secondary afternoon maximum in summer, it is greatest at this season.

The hourly means shown by the daily variation curves in the upper part of Fig. 13 depend much more (roughly in the ratio 3 to 1) on the hours of character-figure 1 than on the hours of character-figure 2. The frequency-distribution of the character-figure 2 throughout the day at each season is shown in the lower half of Fig. 13; they occur most often at 21–22^h in winter, 20–21^h at the equinoxes, and 23–24^h in summer. The minimum frequency occurs, at all seasons, at 10–11^h. There is a marked secondary maximum of frequency (of hours of character-figure 2) in summer at 17–18^h. At all hours of the day except between 15^h and 16^h there are more hours of character 2 at the equinoxes than in the other two seasons.

The data were also grouped according to sunspot epoch. It appeared that the maximum mean hourly figure occurs at 0–1^h throughout the sunspot cycle, but that the minimum hourly figure occurs earlier (8^h or 8–9^h) at sunspot-maximum than at sunspot-minimum (10^h). The daily range of mean hourly figures increases from 0·395 at sunspot-minimum
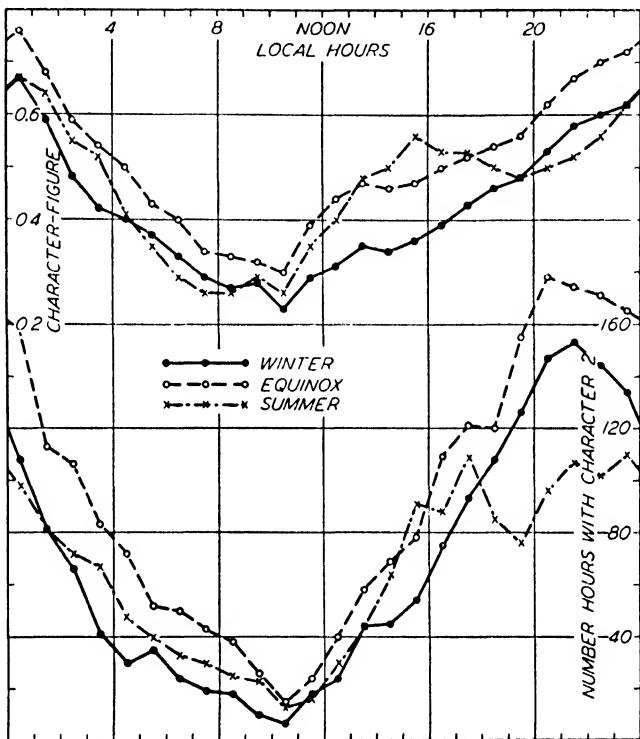
FIG. 13. Above: the daily variation of the hourly magnetic character-figures for Kew observatory, 1913-23. Below: the daily variation of the number of hours of character-figure 2. (After J. M. Stagg)

to 0·445 at sunspot-maximum: these are 110 and 74 per cent. of the respective average character-figure for the two groups of years. The secondary afternoon maximum in summer is better developed at sunspot-maximum than at sunspot-minimum, particularly when years of descending sunspot numbers are compared. The frequency of hours 2 has its minimum at 10–11^h throughout the sunspot cycle.

The increase of sunspottedness in maximum years affects the daily variation of mean hourly character more in winter than in the other seasons.

A complete survey of the daily variation of the magnetic activity is not yet possible, since only a few observatories have discussed this aspect of their records.

J. M. Stagg [42] collected data for the daily variation of magnetic activity at seven stations in the northern hemisphere and three stations in the Antarctic (Cape Evans, Cape Denison, and the Gauss station), by using either hourly character figures, or ranges (differences between maximum and minimum) for hourly intervals. For high-latitude
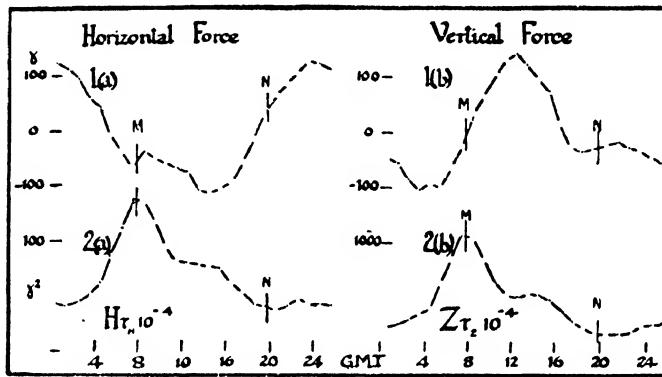
FIG. 14. The disturbance-daily variation (above) and the daily variation of hourly ranges (below) in H and Z at Fort Rae. (After J. M. Stagg)

stations, the regular daily disturbance variation S_D is very large; for instance, at Fort Rae its amplitude exceeds 400γ in the average of five disturbed days per month, even in the magnetically quiet Polar Year 1932/3. The contribution of S_D to the hourly ranges used as an index of disturbance 'might be considered overwhelming in comparison with the contribution of the irregular perturbations' discussed here, but Fig. 14 shows that the facts are quite different: the hourly range data for H and Z (shown below) cannot be regarded simply as the hourly change due to the corresponding S_D variations (shown above in Fig. 14). Fig. 15 shows that the daily variation of activity (measured in various ways) is controlled by local time over the whole polar region covered by the stations, which extend to about 35° from the magnetic axis pole. The form of this variation changes with latitude and season, and with the general level of activity.

Beyond a distance of 20° from the magnetic pole (that is, at geomagnetic latitudes lower than 70°), the daily variation of a is mainly a 24-hour wave with the maximum invariably in the evening; the

maximum becomes later, in this region, as we go polewards, the retardation being at the rate of 1 hour per 5° increase in latitude, till the maximum falls at midnight at 70° (Fig. 15). Between 70° and 80° the daily variation of the activity ceases to have always one dominant maximum, and in this zone its form depends entirely on the season and on the general level of activity (Fig. 16). In still higher latitudes, 80° to 90°, the daily variation has again one dominant maximum, which is always in the forenoon. The highest level of general magnetic activity is highly localized at about 20° distance from the pole (that is, near the auroral zone).

Fig. 17 indicates the special conditions in the region lying within 12° of the magnetic axis pole: it is continually in daylight during summer, and continually in darkness in winter. The main features of the daily variation of activity and of its average intensity are also summarized in this diagram by Stagg.

11.20. The dependence of the interdiurnal variability U on the initial hour of the day. The interdiurnal variability U, which forms the basis of the u-measure of magnetic activity (§ 6), is the average day-to-day change of the daily mean value of the north component X or of the horizontal component H, reckoned without regard to sign. Its value is found to depend materially on the hour of commencement of the 24-hourly interval for which the daily means are computed. This has been shown by Bartels [19], from a study of the X-data for Potsdam and
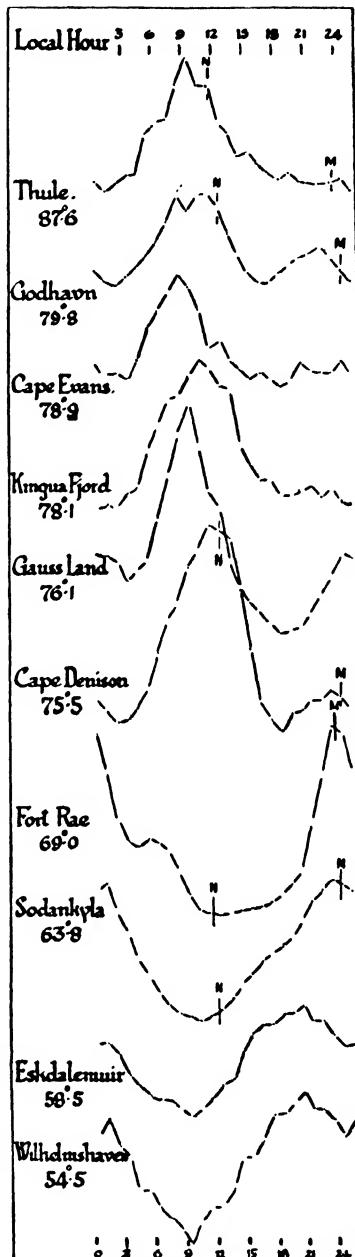
FIG. 15. The daily variation of the hourly magnetic activity, according to local time, at ten polar stations. (After J. M. Stagg)
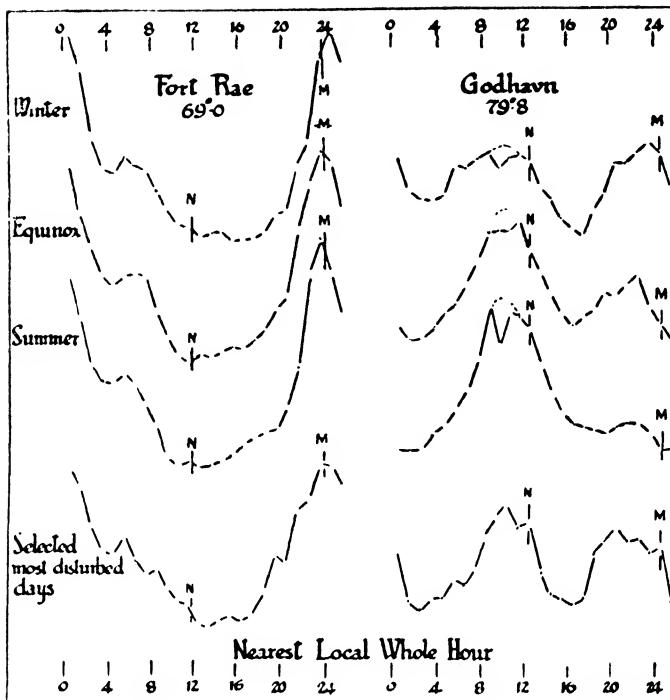
FIG. 16. Indicating how the daily variation of the magnetic activity at Fort Rae and Godhavn depends on the season and on the intensity of magnetic disturbance. (After J. M. Stagg)
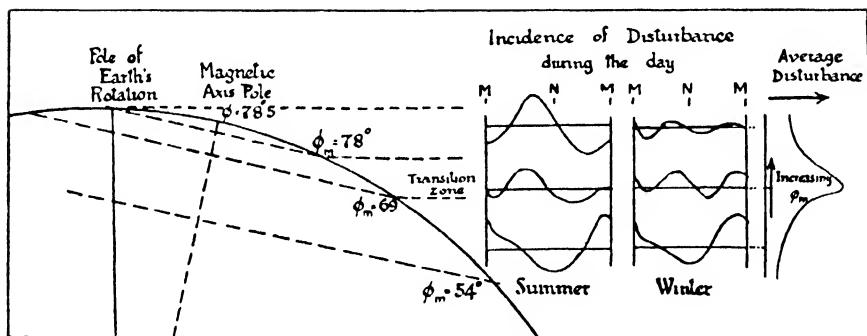
FIG. 17. Illustrating the main features of the daily variation of the magnetic activity in polar regions. (After J. M. Stagg)

Batavia, for the years 1911–20. He considered four different commencing hours, namely, 0, 6, 12, and 18^h G.M.T. A suffix will here be added to U to indicate the commencing hour of the days from which it was calculated, and the mean of U_0, U_6, U_{12}, and U_{18} will be denoted by $\overline{U}$. The following table gives $\overline{U}$ and the differences $U_n - \overline{U}$ for the three seasons and the year: it reveals the surprising fact that the commencing hour which gives the largest value of U varies with the season.

TABLE 6. $\overline{U}$ and $U_n - \overline{U}$ for the north component X at Potsdam and Batavia, 1912–20; the hours n refer to G.M.T.

| | Potsdam | | | | | | Batavia | | | |
| | $U_n - \overline{U}$ for $n =$ | | | | | | $U_n - \overline{U}$ for $n =$ | | | |
$\overline{U}$	0^h	6^h	12^h	18^h	Groups of 4 months	$\overline{U}$	0^h	6^h	12^h	18^h
γ	γ	γ	γ	γ		γ	γ	γ	γ	γ
5·20	−0·52	+0·17	+0·50	−0·14	June solstice	8·17	−0·16	+0·04	+0·29	−0·17
5·61	−0·19	+0·34	+0·06	−0·22	Equinoxes	9·40	+0·16	+0·17	−0·09	−0·24
5·08	−0·03	+0·31	−0·12	−0·16	December solstice	8·05	+0·32	+0·36	−0·26	−0·42
5·30	−0·25	+0·28	+0·15	−0·17	Year	8·54	+0·11	+0·19	−0·02	−0·28

Since the *local* times of Potsdam and Batavia differ by 5·6 hours, the identity of the hours n of maximum or minimum U at the two stations suggests the existence of a 'universal-time' variation of U. Such a variation might depend, for example, on the varying angle between the earth's magnetic axis and the line connecting the sun and the earth. This angle β depends on the angles 11·5° between the earth's rotational and magnetic axes, and 23·5° between the earth's rotational axis and the normal to the ecliptic plane. At the June solstice, β is 35° and 12° respectively when it is noon and midnight at the northerly magnetic pole; at the December solstice, the same holds for the magnetic pole in the southern hemisphere. But, since these two poles are antipodes, their local times differ by 12 hours. Therefore any daily variation depending on β should have opposite phases in June and in December. This is the case in the table, if the means for the year (last line of Table 6) are subtracted; the average of the remaining effect for Potsdam and Batavia is, in γ:

Beginning of day (G.M.T.) . .	0^h	6^h	12^h	18^h
June solstice	−0·27	−0·13	+0·33	+0·07
December solstice . . .	+0·22	+0·10	−0·26	−0·06

For no other station, however, have similar data yet been prepared, so that the facts elsewhere are unknown. But the effect does not

necessarily lessen the value of u as a measure of the magnetic activity, in comparison with other measures, for the simple reason that other measures have not yet been examined in this respect.

11.21. Magnetic effects simultaneous with radio fade-outs. Since systematic observations of radio fade-outs (cf. 10.3) have not been collected prior to 1934, the material at hand is not yet sufficient to make a detailed statistical analysis of their associated magnetic effects. However, it seems certain that these occur only on the sunlit hemisphere. Because of their short duration—an hour or less—they do not raise the estimated magnetic character-figure C very much; for instance, the values of C for the days 1936 April 8 and November 6, when such effects were unusually intense, were only 0·8 and 0·9.

McNish [43], from a first survey, stated provisionally that the ability of a solar eruption to produce a radio fade-out or a magnetic effect does not seem to depend upon its position on the sun's disk. Strong effects were produced on 1936 August 25 by an eruption occurring at a distance of 0·96 of the solar radius from the centre of the disk, while some fade-outs appear to have been produced by eruptions occurring actually on the limb of the sun. The most intense and largest solar eruptions produce both fade-outs and magnetic effects; smaller and less intense eruptions cause only fade-outs, while the least intense but more numerous eruptions produce no conspicuous effects on the earth (see also [39] and [40]).

The number of bright solar eruptions in the light of hydrogen or ionized calcium increases very much from sunspot-minimum to maximum, as is shown by the following table, due to Newton and Barton [5.12]:

TABLE 7

Year	Number of bright solar eruptions of intensity:			Total (including unclassified cases)	Mean sunspot number
	1	2	3		
1934	14	1	0	15	9
1935	137	18	2	161	36
1936	472	134	27	642	80

They also found, using material differing from that of Dellinger [43], that a close relationship exists between radio fadings and solar eruptions, implying a solar agency with approximately the speed of light.

They applied the superposed-epoch method to the international magnetic character figures C, taking as zero-day that of the radio fading;

they found a slight increase of C on the day of the fade-out (in the mean for 37 fadings):

Day . .	-3	-2	-1	0	$+1$	$+2$	$+3$	$+4$	$+5$	$+6$
Mean C .	0·61	0·56	0·69	0·84	0·71	0·65	0·64	0·52	0·67	0·60

11.22. Further statistical aspects of measures of activity.

The high values of the average international magnetic character-figures C for 1930, simultaneous with low values for the u_1-measure, raised a suspicion of a break in the homogeneity of the series of C. However, this discrepancy between C and u_1 may be due to the different aspects of magnetic activity which they characterize. These different aspects may be illustrated as follows:

(a) April 1930 has the highest average monthly character-figure, $C = 1·04$, of the whole series from 1906 to 1934. In the lists of 'Principal magnetic storms' published in *Terrestrial Magnetism*, however, all observatories except Sitka reported 'no storm' for this month. On the other hand, May 1921, famous for two of the most intense magnetic storms ever recorded, has only $C = 0·83$. The u_1-measure gives 52 for April 1930, and 132 for May 1921 (the second highest value for the series 1872–1934). This means that the mean monthly C fails to indicate the conspicuous absence of storms in April 1930, as compared with the outstanding storms in May 1921, while u_1 represents these conditions satisfactorily.

(b) April 1930 had only four days on which $C \leqslant 0·4$, while May 1921 had eleven such days. The measure u_1 fails to indicate the conspicuously low number of quiet days in April 1930, while C is responsive to this feature.

This indicates the following explanation of the above discrepancy: According to C, April 1930 appears as a month of very high activity, namely, with very *few quiet days*; according to u_1, the same month appears as a month of low activity, namely, with *no magnetic storm*.

As an example, let us suppose that month A has 5 very quiet days for which $C = 0·0$, and 25 moderately disturbed days for which $C = 1·0$: and that month B has 15 days for which $C = 0·0$, 5 days with $C = 1·0$, and 10 highly disturbed days for which $C = 2·0$. Both months have the same average, $C = 0·83$, and may seem equally disturbed. Two contrasting opinions can be held, however: either the month A may be judged more disturbed than B, because of the smaller number of quiet days in A; or, vice versa, the month B may be judged

more disturbed than A, because of the great number of highly disturbed days in B. Month A is similar to April 1930; month B is similar to May 1921.

Which of these opinions is more reasonable depends on the circumstances. Imagine two geophysical phenomena P and Q, both depending on geomagnetic activity, but with different sensitivity so that, in P, the highest possible effect of activity is already reached when $C \geqslant 1\cdot0$, while Q is only affected when $C \geqslant 1\cdot5$. Then, as far as P is concerned, month A would appear more disturbed than month B, whereas for Q, month A is quiet and month B is disturbed. Such cases may arise in radio work on the ionosphere (see p. 504).

(c) The characteristic difference between u (or u_1) and C is one of gradation. The increase of daily magnetic activity from $C = 1\cdot7$ (the highest daily figure in April 1930) to $2\cdot0$, judged by u, is much greater than the increase from $C = 0\cdot0$ to $0\cdot3$, but both increases would affect the monthly mean C alike (§ 7).

In conclusion, we may roughly classify solar phenomena according to their effect on the earth's magnetism as follows:

(1) Individual flares of ultra-violet light: these produce brief geomagnetic effects simultaneous with radio fade-outs.

(2) The general change of ionizing wave-radiation in the course of the sunspot cycle; this governs the general intensity of the solar daily variation S_q.

(3) Moderate corpuscular radiation: this produces ordinary aurora and minor magnetic disturbances, and is the main factor governing the daily magnetic character-figure C.

(4) Intense corpuscular radiation: this produces magnetic storms, and aurorae outside the auroral zone. It is the main factor affecting the u-measure of magnetic activity.

11.23. Relation to comets. When Halley's comet passed between the sun and the earth on 1910 May 19, the possibility that the earth might touch or traverse its tail led to many geophysical observations being made. However, no striking phenomenon was observed, nor anything which might not have happened in the ordinary course. The magnetic character-figure for the day was $C = 1\cdot1$.

Maris and Hulburt [21] argued that if the emission which causes terrestrial magnetic storms and aurorae should impinge upon a comet, it might cause changes in the comet. Now many cases of apparently lawless and erratic behaviour of comets have been observed, and after investigating these they concluded that in general the cometary events

(such as sudden brightening, splitting into pieces, emission of a new tail, etc.) followed closely after a strong magnetic disturbance. From a statistical study for 31 comets observed during the years 1848–1927 they inferred that outstanding cometary changes, 28 in number, were observed at an average interval of 5 days after strong magnetic storms.

11.17. Additional notes. Newton [31 a] found that of the 29 bright eruptions of intensity 3+, included in his list of 91 'major' eruptions of intensity 3 or 3+ (p. 193), no less than 21 were followed within 4·0 days by a magnetic storm: and that half of these storms were classed at Greenwich as 'great'. This high degree of association between two infrequent phenomena resembles that found between the biggest sunspots and magnetic storms. He also found a pronounced tendency for the occurrence of a peak in the magnetic activity between 0 and 4 days after a 'major' bright eruption, so long as this was in the *central* part of the sun's disk (between 52° east and 52° west of the central meridian) but not otherwise; eruptions of intensity 2 did not show such a peak even when they were in the central region.

This peak of magnetic activity, associated with bright eruptions in the central region, is specially marked for great storms and for the eruptions of class 3+. During 1935–8, six of the eight 'great' storms that occurred followed 'major' eruptions within two or three days, the mean interval being 1·0 day.

Though bright eruptions are usually found in association with sunspots, these facts do not merely reflect the known statistical relationship between big spots and great storms, because it appears that more than half the spots associated with the 'major' eruptions were *not* big, and that if the spot-size were used as a basis for predicting storms, we should expect only about half the number of storms actually found in association with the 'major' eruptions. The disparity is still greater if we use data relating to 'great' storms.

Newton gives two other pieces of evidence which suggest 'that a major bright eruption is a critical phenomenon in the occurrence of great magnetic storms'. He concludes: 'While it cannot be claimed that a direct relationship between intense bright eruptions and magnetic storms has been well established, the available data seem to show that such a relationship is very probable. (It is anticipated, however, that the occurrence of sequences of smaller magnetic storms during solar minimum may still remain an enigma with respect to any observable solar phenomena.)'

XII

THE 27-DAY RECURRENCE-TENDENCY IN MAGNETIC CONDITIONS

12.1. Results of the superposed-epoch method. In the present chapter another remarkable feature of magnetic disturbance will be considered, namely, the tendency for unusually calm or unusually disturbed conditions to recur after the lapse of one or more intervals of about 27 days.

The simplest method of establishing the existence of the recurrence-tendency, and of determining the average length of the recurrence interval, is the superposed-epoch method, due to Chree [11.29]. He picked out the 5 most disturbed days in each month of a series of years, and wrote down their character-figures in one of a series of columns. In preceding and succeeding columns, in the same horizontal row, he entered the character-figures, in order, for a number of days before and after these selected days. The means for each column were then calculated, thus affording a measure of the average degree of disturbance at different intervals before or after days of marked disturbance.

In his first investigation [1] by this method, for the period 1890–1900, for which no international character-figures were available, he himself had first assigned character-figures 0, 1, or 2 to each day, by inspection of the Kew magnetograms: the 5 selected disturbed days were the days of largest range in H at Kew. In a later paper he considered the years 1906–11, and used international character-figures; this work was extended later, in a joint paper with Stagg [3], to the years 1906–25.

This recurrence tendency seems to have been first noticed by Broun [4] in 1858, and Hornstein [5], Liznar [6], Quetelet, Veeder, Clough, and Harvey [7] also called attention to a 'period' of 26 or 27 days in magnetic or auroral phenomena. But the nature of the tendency was not understood, and even its existence was often doubted or overlooked, until Maunder's rediscovery [8] of it in 1905. Maunder's papers, in which the 27-day time pattern of magnetic activity was first introduced, established the existence of the 27-day recurrence tendency almost beyond question, and drew general attention to its importance and probable significance. Subsequently, Chree removed every doubt on the matter by his superposed-epoch method, which is free from any suspicion of preconceived notions as to the length and significance of the recurrence interval. He applied his method first to the 5 days of

largest character-figure per month as the selected days. The results are shown in Fig. 1; along the abscissae the days are indicated, n denoting the selected days: the tabulations included the 5 previous days and the 30 or 35 following days. The ordinates represent the mean character figures.

The days $n \pm 1$, immediately adjacent to the selected days, are distinctly more disturbed than the average: this is mainly a consequence of the fact that disturbance often lasts for more than 1 day. The significant feature of the diagram, shown by both curves, is the marked rise in the mean character-figure round about the 27th day; it clearly
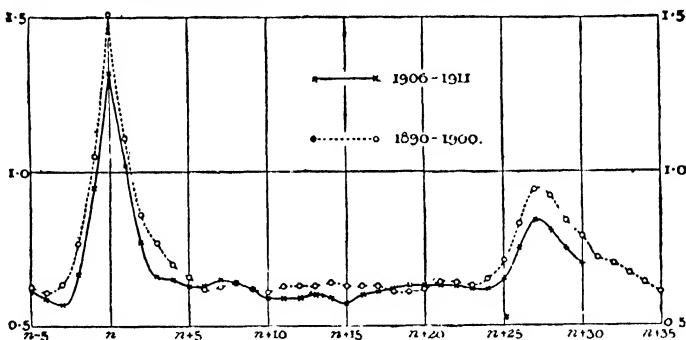
FIG. 1. The 27-day recurrence phenomenon in the daily magnetic character-figures, as shown by Chree's daily character-figures assigned on the basis of the Kew records 1890–1900, and by the international magnetic character-figures, 1906–11. The curves indicate the average daily magnetic character-figures for days preceding and following a set of selected magnetically disturbed days (n)

indicates the tendency for disturbed conditions to recur after this interval. There appears to be no shorter interval of recurrence, and this has been confirmed by further similar investigations.

A series of average character-figures was formed in like manner for days preceding and following selected *quiet* days (5 per month throughout a series of years). The recurrence tendency, with the same interval of about 27 days, was shown equally in this case. The corresponding curve of mean character-figures naturally has its peaks inverted, the selected days, and those about 27 days later, having character-figures below the average (Fig. 2).

Chree showed further that there are corresponding peaks of diminishing height or depth at about 54 and 81 days after the selected days (disturbed or quiet), and also at intervals of about 27, 54, and 81 days before them. This is illustrated both for disturbed and quiet days in Fig. 3; to save labour, on this diagram the portions of the curves
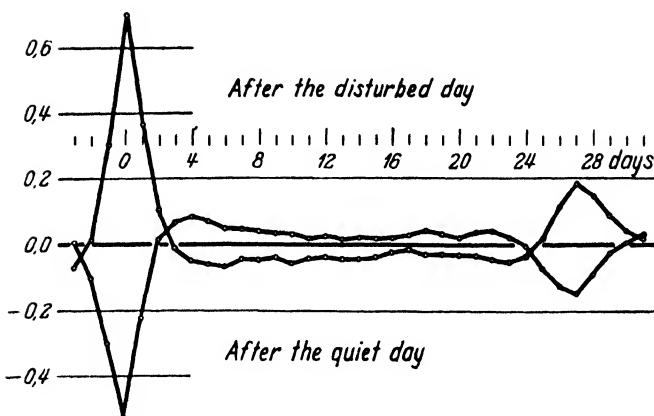
FIG. 2. The average change in the daily magnetic character-figures, indicated by the deviations from the mean, from four days before till 31 days after a set of disturbed or a set of quiet days, 1906–24. (After Chree and Stagg)
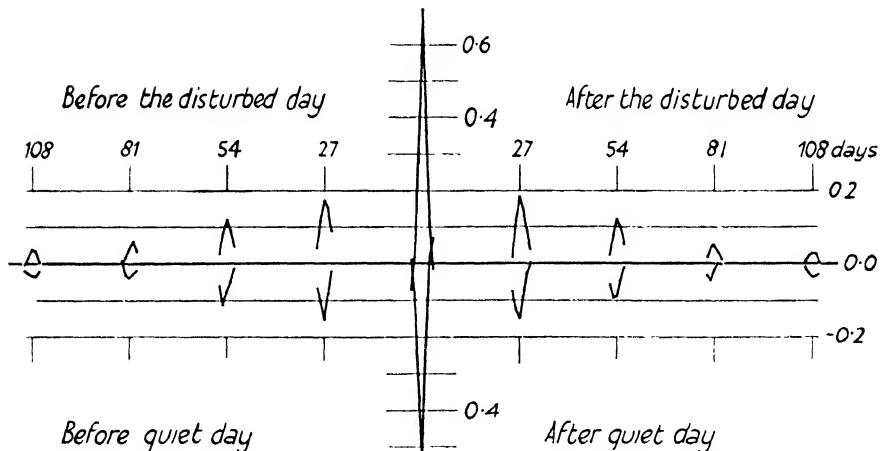
FIG. 3. The average change in the daily magnetic character-figures, indicated by the deviations from the mean, for groups of days around a set of disturbed or a set of quiet days (centre), and around epochs 1, 2, 3, 4, solar rotations (of 27 days each) before and after. (After Chree and Stagg)

between the peaks were not determined. The middle horizontal line corresponds to the mean character-figure, and the upper and lower parallels represent character levels 0·1 and 0·2 greater and less than the average. One of the chief objects in constructing this diagram was to obtain a more exact estimate of the *length* of the recurrence interval, by considering peaks at multiple intervals. The more distant peaks are,

however, too small and irregular to permit a very definite estimate to be made. It may be hoped that this will become possible by the use of more extensive material. Chree's estimate of the interval is 27·0 days; Archenhold [3 a] finds that it varies from 27·0 days to about 27·6 days, being definitely smaller in the years just preceding a sunspot minimum than in the years just after a minimum.

The selected days are all of special character, more quiet or more disturbed than the average. The curves of mean character-figure do not indicate, however, whether the secondary peak in the Chree diagram, 27 days later, is produced by a general small divergence of the 27th days from average conditions, in the same direction as for the primary peak, or by a small proportion of days of marked character, like the selected days, the remainder being average days. Chree investigated this point: the character-figures for the selected 5 most disturbed days per month, in his first paper (for the period 1890–1900), were mainly 2's, but included also many 1's and a few 0's (drawn from quiet years). The percentages of these three character-figures on the selected days, and on the 27th days thereafter, and on all days, were as follows:

Character-figure	Percentage on		
	Selected days	27th days	All days
2	55	22	12
1	40	50	46
0	5	28	42

Thus it appears that the 27th days after disturbed days include a considerable proportion which are quiet, though less than the proportion of quiet days among all days. The proportion of moderately disturbed days (1's) is a little above the average, while that of markedly disturbed days is distinctly above normal. The recurrence tendency consequently does not imply a *certainty* that a disturbed day will be followed 27 days later by another disturbance; the 27th day may, in fact, be quiet, but the chance of this is less than on a day chosen at random, and there is a distinctly greater chance of its being really disturbed.

The mean recurrence interval found by Chree, 27·0 days, is very close to the mean rotation-period of the sun (Ch. V), though there are lunar periods of nearly the same length, namely, the draconitic month of 27·21 days, the sidereal of 27·32 days, and the anomalistic (from one apogee or perigee to the next) of 27·55 days. In view of the other close connexions between terrestrial magnetic disturbance and the sun, there

is little room for doubt that the recurrence tendency is due to the solar rotation.

Stagg [9] applied the superposed-epoch method to his *hourly* character-figures (p. 386), in the hope that they would help towards a closer analysis of such influences as the drift of the average latitude of sunspots towards the equator in the course of the sunspot cycle, and the corresponding change of rotation-period of the mean sunspot zones at different times during the cycle. In the two groups of high and low sunspot latitudes, the average latitudes differ by an amount which would correspond to an increase of the solar rotation period from low to high latitudes by 0·26 days (see 5.4). The strong daily variation of magnetic activity, however, introduced complications. The most favoured epoch for recurrence is approximately 27 days from the originally disturbed hours; it is, however, almost a matter of chance whether a particular disturbance reaches a higher state of development after this interval or at a time differing by about 24 or 48 hours from the complete 27 days.

It may be noted that Chree (in Part 2 of his paper [11.29]) examined also the recurrence tendency in sunspots (mainly due, of course, to the survival of spots throughout a whole rotation or more) by his method. He took as selected days the 5 per month which had the largest 'projected' spot area. For these and for 10 days before and after them, and also for the 25th to the 30th day after, he formed the averages of the daily projected sunspot areas. In the latter group of days he found a repetition of the primary peak on much the same scale, relative to the latter, as that shown for secondary and primary peaks on the diagrams of average magnetic character-figures in Fig. 1. The spot-peaks, however, were much more rounded than those in Figs. 1–3, which is natural in view of the continued visibility of spots, and their slow progress across the sun's disk, during several days.

12.2. The recurrence of geomagnetic storms. Maunder [11.34], by his study of the times of occurrence of magnetic storms, was led to conclude that the magnetic recurrence interval is determined by the period of the solar rotation, and his presentation of the evidence was designed to enforce this conclusion, and was stated in terms of the solar rotation. What he proved, however, is that magnetic storms tend to recur after an interval of about 27·3 days, and the identification of this with the solar rotation period is an independent step which, though highly probable, is a speculation rather than an established fact, like the recurrence tendency itself.

Maunder's material consisted of a list of the principal magnetic disturbances observed at Greenwich during the period 1848–1903 (afterwards extended to 1913). The whole series includes 816 disturbances, that is, about 13 per year. The disturbances were more extreme than those afterwards considered by Chree, which numbered 5 per month, or 60 per year. Those considered by Maunder had been catalogued by Ellis for the period up to 1882, while the later ones consisted of all those for which reproductions of the magnetographs had been given in the Greenwich volumes, provided that their range in declination was at least 20'. The material was thus selected without reference to the investigation of the recurrence tendency.

The list of disturbances gave, among other particulars, the time of their commencement and their duration. To these Maunder added a further two columns, giving the heliographic latitude and longitude of the centre of the sun's disk, that is, of the earth, at the commencement of the storm. The longitude (5.4) was measured from the prime meridian rotating with the adopted mean rotation period 27·275 days: the number of complete rotations made up to the time of the storm was also given, indicating in which rotation-period the storm commenced.

Inspection of the list showed many pairs or series of storms occurring during successive or adjacent rotations, and with the earth in nearly the same heliographic longitude at their commencement, or, otherwise stated, with nearly the same meridian on the sun directed towards the earth. The agreement between the heliographic longitudes may, however, be regarded as merely one way, not necessarily possessing any intrinsic physical significance, of indicating that the recurrence interval between the storms in question was nearly 27·27 days. (Owing to the varying orbital motion of the earth, the interval corresponding to successive passages of a particular meridian of the sun across the centre of the disk changes slightly in the course of the year—Carrington's zero meridian returns to the centre after 27·20 days at the end of June, but after 27·33 days at the end of December, cf. p. 169.)

The number of cases of recurrence of storms after a lapse of one or more intervals of about 27 days was much greater than could be attributed to chance. In the period 1882–1903, which included 276 catalogued storms, there were 49 pairs of consecutive rotations of the sun in each of which one and only one storm occurred. Of these 49 pairs, 18 gave the second disturbance as corresponding to the same solar longitude, within 10°, as the first: the chance of this occurring fortuitously in 18 out of 49 pairs is only 1 in 20 thousand millions. Again,

in the same 276 storms there were 36 sets in which a disturbance in one rotation was followed in the next by another when nearly the same heliographic meridian passed through the earth; in 8 of these 36 cases there was a further similar repetition in the third consecutive rotation, in 4 more the repetitions were extended to the fourth rotation, and in one other they continued through 6 rotations in all. In addition, there were 16 cases in which the interval was not one rotation but two. Nearly one-half of the 276 storms were included in one or other of such sequences, and in a large proportion of cases no other disturbances occurred within the single or double rotation interval. When higher multiples of a rotation period than two were included, more than three-quarters of the 276 storms could be grouped in some shorter or longer sequence.

One very clear case may be cited of a sequence of storms some of which, if associated at all with the passage of a particular region of the sun across (or near) the middle of the disk, occurred either before spot-activity appeared in the region, or after it had ceased. The sequence was as follows (XXXII of Maunder's list):

No. of disturbance	Date of storm 1889	No. of rotation	Class	Coordinates of the centre of the sun's disk	
				Longitude	Latitude
104	July 17	478	Active	57°	5°
105	Aug. 13	479	Moderate	60°	7°
106	Sept. 8/11	480	,,	66°	7°
108	Oct. 5/6	481	,,	73°	6°
111	Nov. 1/2	482	Active	83°	4°
112	Nov. 26/29	483	,,	109°	1°

During the five months over which this sequence extended, only three other disturbances occurred (Nos. 107, 109, 110), and these were just half a rotation different from them in time, so that they occurred when the side of the sun visible at the time of the above disturbances was turned away from the earth. The year was 1889, one of sunspot-minimum. When the first of the six storms occurred, the largest group of the year (No. 2090, of area 501 millionths) was crossing the disk for the second time as group No. 2092, and a new intermittent group, which in the following rotation became the third largest group of the year, had just appeared in the same latitude as No. 2092, but preceding it in longitude by 48°: it is not clear whether the disturbance No. 104 is to be associated with the considerable spot-groups No. 2092, or the nascent group No. 2093, or with both. The two groups returned in the next rotation as Nos. 2098 and 2099, and the latter was represented

for a couple of days in the next following rotation by a very small group, but the magnetic disturbances were renewed (the last two being 'active') for three rotations after the disappearance of the last feeble remnant of the spot groups.

The heliographic longitudes of the centre of the sun's disk, not only for the commencement but also throughout the duration of each of the 276 storms catalogued for the period 1887–1903, are indicated in Fig. 4, where each disturbance is represented by a straight line (the length of which represents the duration of the disturbance), at a level corresponding to the rotation interval in which the storm occurred; the levels of successive rotations are drawn at equal intervals on the diagram (cf. Fig. 5.6, in which a similar arrangement was adopted to display the longitudes of sunspots). Recurrences of storms after one or more solar rotations are indicated on this diagram by the lines falling nearly above one another, and many examples of this are obvious. When the lines fall nearly but not quite above one another, the recurrence interval differs slightly from 27·3 days. Some sequences show a longer recurrence interval than this, others a shorter one; Maunder found that the recurrence intervals vary over about as wide a range as the values of the solar rotation period derived from different sunspots, whether in the same or different latitudes.

Other features in which the recurrences of magnetic storms recall analogous characteristics of sunspots may be cited. The first is the total period covered by a simple sequence of recurrences; out of the 36 sequences mentioned above, 23 were simple repetitions, 8 double, 4 treble, and in 1 there were 5 recurrences of the first storm, the sequences extending over 6 rotations in all; this proportion 'corresponds very well with the longevity of spot-groups of the more important and stable kind'.

A second similarity is that of intermittence; this has been illustrated in 5.6 as regards sunspots. It may be illustrated for magnetic storms by the following example, covering a period of 59 weeks, during which there were in all 16 catalogued disturbances: 9 of these storms synchronized with the return of one particular meridian to (near) the centre of the sun's disk, and 3 other storms synchronized with the return of another meridian. Many similar but less extended sequences were cited by Maunder in Table VI of his paper, and they were far more numerous than could be attributed to chance, though some of the recurrences may be merely accidental.

A third similarity is that just as, over a number of years, sunspots
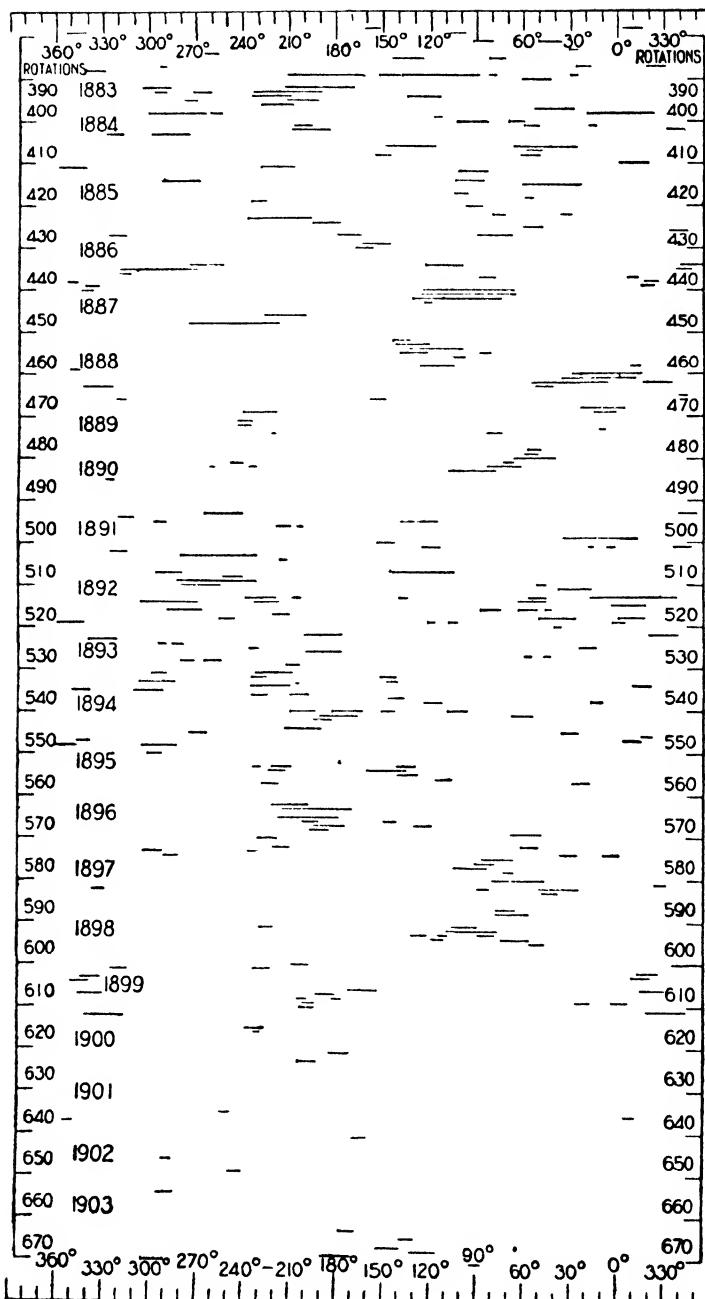
FIG. 4. The distribution of magnetic disturbances, 1882–1903, according to the heliographic longitude of the centre of the sun's disk at the time of their commencement. (After E. W. Maunder)

tend to occur in definite longitudes on the sun, so also the solar meridians passing through the earth during magnetic storms tend to be in definite longitudes. This is well indicated by Fig. 4, which may be likened in this respect to Fig. 5.13, p. 181.

12.3. Lack of recurrence for the most intense storms. Chree established the recurrence tendency for magnetically quiet or moderately active days (60 per year of either kind), as had already been done by Maunder for more notably disturbed days (13 days per year). Later Greaves and Newton [10] examined whether the tendency persists in regard to still more intense storms. They considered 60 great storms, and 343 lesser storms, in the period 1874–1927 (already discussed in 11.16); thus the great storms numbered about 1 per year, and the lesser storms about 7 per year, on the average. The lesser storms were subdivided into four nearly equal smaller groups, according to the value of the mean of the ranges in H, Z, and D (all in force units); the limiting ranges for the four groups were $\geqslant 180\gamma$, 150 to 179γ, 130 to 149γ, and $\leqslant 130\gamma$. Any day including the whole or part of any of the 403 storms was called a storm day; the number of these exceeded 403, because the storms generally extended over 2 or more days. For each of the five groups of storms, the percentage of cases in which the nth succeeding day was a storm day was determined; the values of n considered ranged from $n = 1$ to $n = 95$.

It appeared (Fig. 5) that for the 60 great storms there was no special tendency for the days round about $n = 27$ (or 54 or 81) to be storm days; also for the next group of 88 less intense storms, the tendency, if present at all, was extremely slight. For the three remaining groups a decided peak around $n = 27$ was visible, and for the two groups of least intense storms there was also a peak around $n = 54$, though none around $n = 81$.

Greaves and Newton found 20 cases in which sequences of 3 or more of their 403 storms followed each other at intervals between 25 and 29 days; 75 storms were included in these sequences, but only 3 of them were 'great' storms—a proportion, 1 in 25, much less than that (60 in 403, or about 1 in 7) of these great storms to the whole number.

It thus appears that the recurrence tendency is characteristic of storms of smaller range, and of minor magnetic activity or calm. This fact in no way conflicts with the conclusion drawn by Maunder from the recurrence tendency, namely, that magnetic storms are due to the action of laterally limited streams proceeding from restricted solar areas. But it suggests the additional conclusion that the more intense

storms are due to a short-lived solar disturbance, which does not usually survive a solar rotation.

At first sight it seems paradoxical that the larger storms, which are precisely those that are most associated with sunspots (11.16), should not exhibit the recurrence tendency, since the larger spots show a
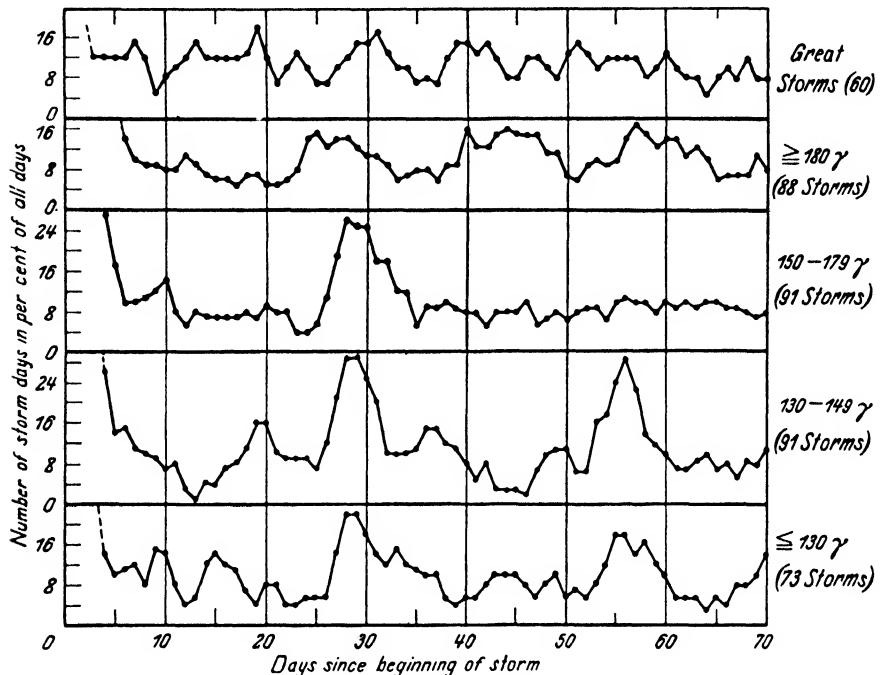
FIG. 5. The average frequency of magnetic storms during the first 70 days after the outbreak of selected storms, shown separately for five groups of initial storms of different intensity. (After Greaves and Newton)

marked tendency to persist for more than one rotation, and therefore to recur near the sun's centre. But the paradox lessens when it is remembered that the sunspots are not themselves the necessary cause of storms (as shown by the occasional occurrence of a storm during a spotless period), but are only associated with the cause. The latter may rise to abnormal intensity for a short period, usually in the presence of a large spot, but subside to insignificance during part of the lifetime of the spot.

In one respect the facts for the large and the moderate storms are supplementary in indicating that magnetic disturbance is connected with localized activity on the sun's surface; the recurrence tendency

suggests this for the smaller storms, and the correlation with sunspot data points to the same conclusions in the case of the larger storms.

12.4. The 27-day time-pattern. A convenient modification of Maunder's time-pattern was introduced by Bartels [11], for the study of recurrence phenomena. All days are arranged in rows of 27 days, called 'rotations'; the date of the first day in each rotation is indicated at the left; in addition, the rotations are numbered, No. 1 beginning with January 11, 1906, because the original international scheme of magnetic characterization of Greenwich days started in 1906. Later this rotation was renumbered 1001. On squared paper, one square is assigned to each day, and one of seven symbols is inserted in it, with various degrees of blackening to correspond with various degrees of intensity in the phenomenon studied (starting with a blank for group 0, and ending with a completely black square for group 6). The assignment of symbols to the seven groups of days was chosen as follows, for the international magnetic character-figures C, and for the relative sunspot numbers N for the central circle of the sun (of half-diameter):

Group .	0	1	2	3	4	5	6
Symbol .	White square	Small dot	Small ring	Larger ring	Black circle	Black octagon	Black square
C . .	0·0 to 0·1	0·2 to 0·3	0·4 to 0·7	0·8 to 1·0	1·1 to 1·3	1·4 to 1·6	1·7 to 2·0
N . .	0	1–8	9–18	19–29	30–46	47–61	over 62

Each successive row begins 27 days later than the preceding one; at the end of each row the first 6 days of the next row are repeated in order to emphasize the continuity of the series. Thus actually there are 33 days in each row, of which the first 27, for reference, are numbered 1 to 27. Fig. 6 gives an example (in which there are only six groups, 0 to 5, instead of seven), and Fig. 7 another.

12.5. The time-pattern for international magnetic character-figures 1906–33. The pattern has the following characteristic features:

(*a*) Several like symbols often follow each other in the horizontal direction. This corresponds to the fact that quiet or disturbed conditions do not occur at random, but persist for several days on end (this property is called 'positive conservation', and is discussed further in 16.28).

(*b*) The recurrence tendency is clearly indicated by *vertical* columns of symbols of the same kind (*sequences*).

(*c*) Not a single rotation is entirely quiet ($C < 0.8$) or entirely disturbed; moreover, pronounced sequences of quiet days persist near

sunspot-maxima, and sequences of disturbed days persist near sunspot-minima.

(*d*) A slant in the sequence-columns indicates recurrences after intervals shorter or longer than 27·0 days. Such slants are frequent, but not systematic.
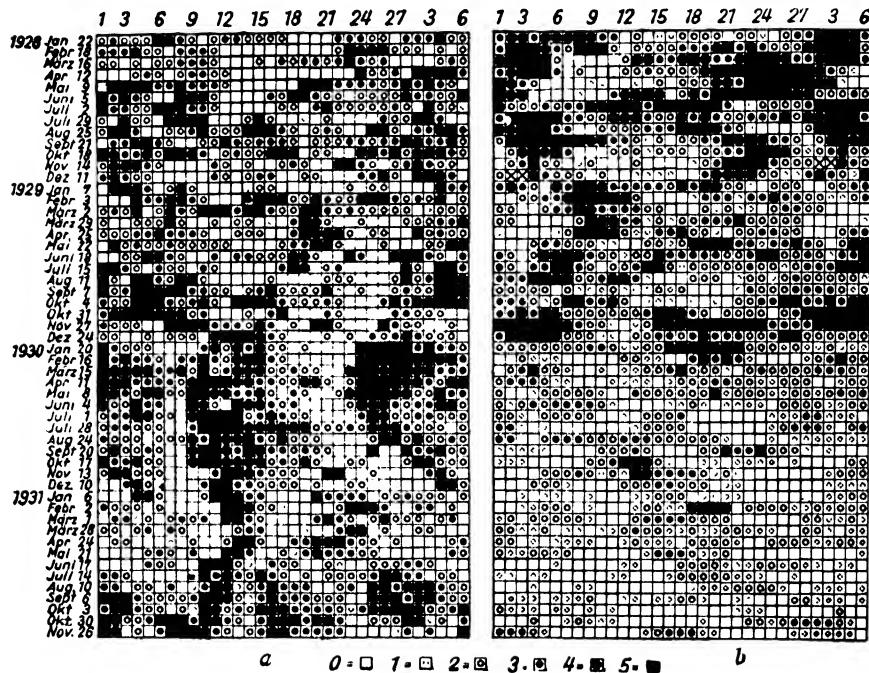
Fig. 6. Time-patterns, for successive rows of 27 days, illustrating the 27-day recurrence tendency for the daily magnetic activity (left, *a*) and the daily sunspottedness (right, *b*). The date for the first day in each row is indicated at the left (§ 4)

(*e*) Some great storms appear in isolation, that is, within an otherwise quiet vertical sequence (see § 4). Other great storms, however, appear in, or lead, pronounced vertical sequences.

(*f*) Especially long sequences seem to occur at the end of a sunspot-maximum, when the spots are near the sun's equator; the first disturbances of the new cycle are more irregular, giving those parts of the diagram a spotty appearance. The transition from the old to the new cycle is well marked in 1914 and 1923. The fine sequence in the first half of 1923 (day 18), with its sharp recurrences after 27 days, may be attributed to the old cycle (Fig. 7); it is remarkable because it persists through times (in February and May) when, for several weeks in succession, not a single sunspot was visible. Disturbed sequences also

persisted throughout 1913, although 'no year since 1810 has been so barren of sunspots as 1913'.

(*g*) Though often two or more sequences run simultaneously, they do not divide the 27-day interval into regular subdivisions.

(*h*) The determination of the lengths of sequences is somewhat uncertain, but the existence of very long sequences is obvious. Outstanding examples of disturbed sequences ($C > 0.7$) are

>   1910 July 24 to 1911 June 13, 13 rotations (already discussed by Angenheister [12]);
>   1921 December 11 to 1922 November 27, 14 rotations;
>   1929 December 9 to 1931 March 13, 17 rotations;
>   1931 June 26 to 1932 August 4, 16 rotations.

Examples of quiet sequences ($C < 0.6$) are

>   1912 November 7 to 1913 December 16, 16 rotations;
>   1923 May 16 to 1924 June 28, 15 rotations;
>   1926 August 28 to 1927 August 14, 14 rotations.

These and other long sequences persist regardless of season.

These sequences are much longer than the average lifetime of sunspots. In her catalogue of recurrent groups of sunspots, 1874–1906, Mrs. Maunder [13] listed 624 in all, of which 468 appeared in two rotations, 118 in three, 25 in four, 12 in five, and in one instance only did a sunspot group survive to be seen in a sixth apparition.

The longer duration of magnetic than of sunspot sequences may, of course, in some few cases be caused by a more or less accidental continuation of a source in the northern solar hemisphere by one in the other, but it is perhaps more relevant to mention, in this respect, the longer life of faculae. According to observations at Greenwich 1874–1917 (5.11):

'The duration of faculae connected with a spot-group is on the average at least three times the length of the accompanying spots, but the proportion is a very variable one. . . . The faculae of large spot-groups as compared with those of smaller groups do not last proportionally as long. . . . The faculae frequently act as a connecting link between successive spot-disturbances in the same region. Near the maximum of the solar cycle, some of these centres have been traced without intermission for over six months and a few for nearly a year. . . . The groups of faculae considered as entities are more stable than groups of spots. . . . Sunspots and faculae are very closely connected; there are no spots without faculae, and no extensive areas of faculae without spots.'

12.6. Comparison of the 27-day time-patterns in magnetic and solar activity. Fig. 7 gives 27-day time-patterns for the years 1923–33 (groups as indicated in § 4). Patterns (not reproduced here) were also drawn for various solar character figures (p. 193). The magnetic and solar time-patterns may be compared as follows:

(*a*) The 11-year cycle is much more pronounced in the solar patterns.

(*b*) The sequences in the solar patterns are shorter; the patterns for the different solar phenomena resemble each other closely; but *no such resemblance appears between the pattern for C and any of the solar patterns*, even if the possibility of a general lag of one or more days is taken into account. A statistical calculation confirms this impression: for 1928–30, the correlation coefficient between the daily group indices for relative sunspot numbers and bright hydrogen lines is $+0\cdot74$, whereas it is only $-0\cdot08$ between C and the sunspot numbers. The 36 monthly means, 1928–30, give correlation coefficients as high as $+0\cdot86$, $+0\cdot75$, and $+0\cdot86$ between the sunspot numbers and the indices for bright and dark hydrogen lines and calcium flocculi. The high correlation between the solar patterns removes any hope that, had a more extended programme of such solar observations as are here considered been in progress during the many decades for which sunspot numbers are available, it would have afforded a measure of solar activity better correlated with C than the sunspot numbers are.

12.7. Solar M-regions. The 'magnetically effective' regions of the sun, suggested by the geomagnetic time-pattern, have been called '*M-regions*' by Bartels [11]. There is, of course, no suggestion that they exert a magnetic influence by their own magnetic field, which is likely to be insignificant at the distance of the earth. But they may be pictured as restricted areas responsible for geomagnetic disturbances presumably by emitting corpuscular radiation. They appear to be more long-lived than sunspots. The identification of the M-regions with sunspots or other solar phenomena is sometimes possible, while in many cases the M-regions lead, so to speak, an independent life.

In this way geomagnetic observations supplement the direct physical observations of the sun, by revealing persistent solar influences, indicated by strong 27-day recurrences. Fig. 8 shows the 27-day recurrence tendency in the types of daily variations on individual days at Aso (see 7.9), after Hasegawa [7.21 *b*].

12.8. Twenty-seven-day recurrences in other terrestrial phenomena. Because of the strong connexion between geomagnetic activity, aurorae, and earth currents, it is natural that the 27-day
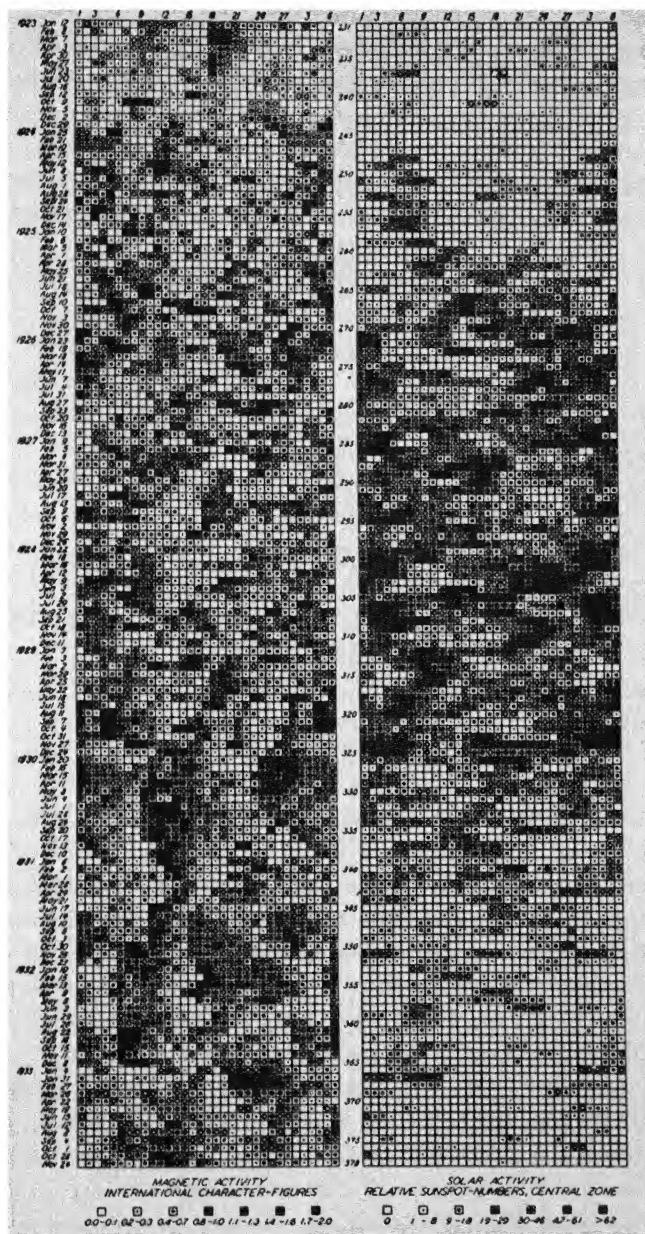
Fɪɢ. 7. Twenty-seven-day recurrence time-patterns for geomagnetic activity (left) and solar activity (right), 1923–33 (§ 4)

recurrence tendency appears also in these phenomena. Thus Sverdrup [G 98], from the auroral character-figures of the Maud expedition, found a recurrence interval of between 27 and 28 days, just as Fritz did from earlier auroral observations; and Peters and Ennis [14] found a recurrence interval of 27 days from earth-current records.

The character of radio communication is measured by the ratio of uncommercial time (time lost to traffic) to total time. Skellett [14a],
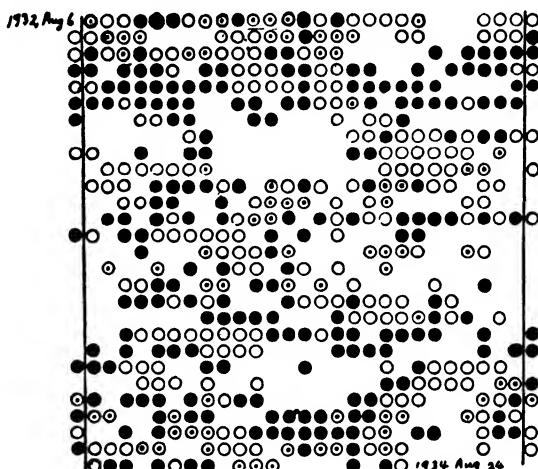
FIG. 8. A 27-day time-pattern illustrating the occurrence of days with different types of daily variation of horizontal magnetic intensity at Aso, 1932 August 6 to 1934 August 24. Black circle, equatorial type E; white circle, polar type P; white circle with dot, intermediate type P'. After Hasegawa [7.21 b]

Berkner and McNish [11.15] found, for short-wave transatlantic telephone circuits, that this ratio increases with the magnetic character-figure C; the 27-day pattern in the 'radio-character' (using numbers 0, 1, 2, 3, for proportions of uncommercial time subdivided according to the limits 25, 25–49, 50–74, 75–100 per cent.) is so marked in the years 1929–32, and so similar to the pattern in C, that Skellett holds that predictions of radio-traffic conditions on the basis of the 27-day recurrence tendency are justified (just as the Norwegian auroral observers always look out for another aurora about 27 days after a display).

Schindelhauer [16] found a 27-day recurrence tendency in wireless atmospherics (statics), but a definite relation to magnetic character-figures could not be established.

12.9. The geometry and width of solar streams of corpuscles. The lateral extent of the solar corpuscular streams which Maunder

suggested as being the cause of magnetic disturbance can be estimated in various ways. Hulburt [G 16] distinguished between two extreme views: one assumes that the solar emission is in a narrow beam (as in the schematic Fig. 9), as from a searchlight or a shotgun, the other that the emission is in a wide flare as from a volcano.

(a) Chapman [17] discussed the geometry of solar streams of corpuscles (24.3). Consider first a point-source at a great distance; if each particle has almost radial motion, the form of the stream (composed of all the particles successively emitted from the same point) will rotate
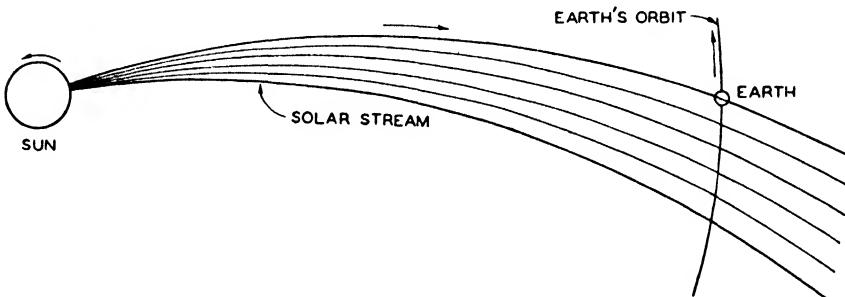
FIG. 9. Sketch illustrating a solar stream of particles moving with a speed corresponding to a time of travel of 36 hours from the sun to the earth

with the sun much like water from a rotating garden-hose. The stream will form an angle χ' with the direction of the radius; at the distance r from the sun,

$$\tan \chi' = -\Omega r/v,$$

where Ω is the angular velocity of the sun's surface, and v the linear radial velocity of an individual particle (assumed uniform). Taking $r = 1\cdot5 \times 10^{13}$ cm. (the distance from the sun to the earth), and

$$\Omega = 2\pi/25\cdot4 \text{ days} = 2\cdot85 \times 10^{-6} \text{ sec.}^{-1},$$

we find that $\tan \chi' = -4\cdot3 \times 10^{7}/v$. The following table gives, for various values of v, the angle χ', the time Δt of passage from the sun to the earth, and the (slightly larger) time-difference $(t' - t_c)$ between the time t_c when the point-source crosses the sun's central meridian, and the time t' of the impact on the earth; it also gives the longitude (λ, east of the central meridian) of the point-source at the time of emission for those particles which actually impinge on the earth.

v (cm./sec.)	10^8	$1\cdot6 \times 10^8$	5×10^8	10^9	10^{10}
$-\chi'$	23°	15°	5°	2·3°	0·2°
Δt (days)	1·7	1·1	0·35	0·17	0·017
$t' - t_c$ (days)	1·9	1·2	0·38	0·19	0·019
λ	24°	15°	5°	2·4°	0·2°

The stream would overtake the earth on the *post meridiem* side; if $\Delta t = 1.1$ day, $-\chi' = 15°$, so that the stream, if the (large) deflexion by the earth's magnetic field were disregarded, would first meet the earth on the 5 p.m. meridian. It would sweep across the earth in 32 seconds (this is calculated from the angular diameter δ of the earth as viewed from the sun—which is 0.85×10^{-4} in angular measure—the time being $\delta/2\pi$ times the rotation period, taken as 27.3 days).

If the earth is enveloped within the stream for one day, the angular width of the stream may be estimated as $360°/27.3 = 13.2°$. The great storm of 1909 September 25 was simultaneous with a spot-group near the central meridian, and lasted only 10 hours; this would correspond to a stream-width of only 5.5°. But often disturbances last longer than a day, and the width thus estimated will be mostly of the order 25°.

(*b*) 'Faculae frequently appear in streaks roughly at right angles to the direction of the sun's rotation, and have a strong tendency to spread from a spot-disturbance for several degrees in latitude. This feature contrasts with the spots themselves, which invariably stream out in longitude, whilst there is little trend in the direction of latitude' [13 a].

This solar evidence leaves the question open whether the cross-section of the stream near the earth is likely to be circular or, if not, whether it will be elongated in or perpendicular to the plane of the ecliptic.

(*c*) Consider a conical stream of semi-angle α, emitted so that its axis is inclined at the angle β to the sun's equator; if the emission is radial, the emitting-point must be in heliographic latitude β, but this is not necessary. The stream will sweep over any distant object of which the heliographic latitude lies between $\beta \pm \alpha$; the heliographic latitude δ' of the earth must probably lie between these limits, if a magnetic storm is to occur. The ecliptic is inclined at 7.3° to the sun's equator, and the earth crosses the sun's equator on about June 5 and December 5, i.e. near the solstices, and attains maximum northerly and southerly heliographic latitudes 7.3° on about September 5 and March 5. Near the former dates, when $\delta' = 0$, streams radially emitted from areas in latitude $\pm 12.5°$ must be of angular width not less than 25° in order to envelop the earth. When $\delta' = 7.3°$ N., the angular breadth need only be 11° for streams radially emitted from latitude 12.5° N., but for streams from an equal southerly latitude it must be at least 40°. It would therefore seem likely that in September magnetic disturbance is commonly due to northern solar emitting areas, and in March to southern areas.

A test of this assumption, made by Bartels [18], is based on the fact

that the two hemispheres of the sun vary, to a certain degree, independently in the 11-year cycle (p. 179). For instance, for years in succession the southern hemisphere may show more sunspots than the northern hemisphere (as in the years 1883–9, 1895–1900, 1907–12); in such years we should expect higher magnetic activity in March than in September, and conversely when the northern hemisphere is the more spotted.

Suitably arranged observational material was available in Maunder's table [19] for the mean daily areas of faculae, corrected for foreshortening and expressed in millionths of the sun's visible disk, for each month of the 30 years 1886–1915. From the 180 months February, March, April, August, September, and October in these years, the individual months were selected in which the total facular area was more than 300, and in which the area of faculae on the northern (or southern) hemisphere exceeded the area on the other hemisphere at least in the ratio 3 : 2. These selected months were then divided into two groups: (1) the 'favourable' group, containing all those single months February, March, April in which faculae were more frequent in the southern hemisphere, and all those single months August, September, October, in which faculae were more frequent in the northern hemisphere; and (2) the 'unfavourable' group containing those months in which the other hemisphere was preferred. The results are as follows:

	Favourable	Unfavourable
Number of months in group	29	33
Mean areas of faculae on the solar disk . .	1,180	1,158
Mean areas on the same side of the sun's equator as the earth	835	380
Mean areas on the other side of the sun's equator	345	778
Mean magnetic activity u_1	50·2	52·3

The effect looked for is conspicuously absent, storms in 'unfavourable' months being at least as frequent as in 'favourable' ones.

(d) This failure makes it difficult to explain the pronounced annual variation of magnetic activity as due to the solar axis not being perpendicular to the ecliptic. This explanation would (crudely stated) imply that streams of angular width of 11° and more should be more than twice as numerous as those of which the angular width exceeds 25°; for at the equinoxes all the former, from one zone only (if we may ignore the few, of breadth greater than 40°, from the other zone) will sweep across the earth, while near the solstices streams from both zones equally will traverse the earth, but only if the width is 25° or greater;

and the conjunctions are more frequent near the equinoxes than near the solstices.

(*e*) If the solar *M*-regions (§ 6) responsible for magnetic disturbance on the earth should follow the drift of sunspots from high to low solar latitudes during the cycle, the longer rotation-period of higher solar latitudes should be reflected in a lengthening of the 27-day recurrence interval (p. 399).

Moreover, the fact that 27-day recurrence sequences are longer at the end of the solar cycle (§ 5 (*f*)) may be taken as indicating that at that time the *M*-regions are nearer the equator and aim their emissions more accurately toward the earth, so that the streams sweep across the earth at every rotation during their lifetime, while at the beginning of the cycle they are emitted at such an angle to the ecliptic that they reach the earth only occasionally. But, of course, an alternative interpretation is that *M*-regions live longer at the end of the solar cycle.

(*f*) The evidence as a whole suggests a width of from 10° to 30° for the solar streams which cause magnetic storms.

12.10. Indication of 30-day recurrence in great storms. The twelve largest magnetic storms in the years 1890–1921, for which at Potsdam the sum of the ranges in the three force components exceeded $1,000\,\gamma$, can be arranged in three groups of recurrences separated by intervals which are multiples of 29·97 days (Schmidt [20]). Taking the noon of 1858 November 16 (Julian day number 2,400,000) as the origin of a numbering of days, the times of commencements of these storms· are given by $29\!\cdot\!97n+r$, where n is an integer and r, in the three groups, is 4·8, 23·7, 16·8. Angenheister [12] found indications of a 30-day recurrence interval in days for which $C = 1\!\cdot\!8$ to 2·0. Also Pollak [16.21], in his periodogram of the magnetic character-figures from 1906 to 1926, found a high amplitude for persistent waves of 29·9 and 30·1 days. Schmidt [20] takes the 30-day period as an indication of a deeper layer of the sun, rotating more slowly than the surface; these eruption centres might persist for years, but act rarely.

However, these periods of length about 30 days still appear doubtful, because rigorous statistical tests show that the evidence for them is not yet sufficient: the arguments of Schmidt [20] are especially convincing in that respect. The reader is referred to 16.29, 16.30, and 16.33, where the general test methods are explained, and applications to the 27-day recurrence tendency are described.

XIII

EARTH-CURRENTS

13.1. Introduction. Variable electrical currents in the earth's crust were first observed when telegraphy along extended lines was introduced. In the circuit formed by the line, grounded at the ends, and the earth, irregular currents of much higher intensity than the artificial currents used for telegraphic communication appeared occasionally and obliterated the signals. Such 'earth-current storms' occur at the same time as magnetic storms and aurorae, and therefore share several characteristics such as world-wide appearance, relations to sunspots, 27-day recurrence tendency, higher frequency at the equinoxes, and so on. This similarity points to a close physical relationship between the two phenomena, and is the main reason why these 'telluric currents' are discussed here along with geomagnetism.

Burbank [2], in a bibliography of the early literature, ascribes to Barlow [1] the first observations 'On the spontaneous electric currents observed in the wires of the electric telegraph', made in 1847 on the English telegraph lines. In the same bibliography we find the characteristic title of a German book by Clement [3] on 'The great northern light in the night before the 29th of August, 1859, and the confusion of the telegraphy in North America and Europe'; aurora was seen in latitudes as low as 14° (north) on the Atlantic, and on some telegraphic lines in France the greatest effects observed were equivalent to 800 volts for a distance of 600 km. Natural potential differences of several hundred volts also occurred in Norwegian cables during the great magnetic storm of 1938 April 16 (p. 329).

Systematic observations of earth-currents are best made by two more or less independent steps. (*a*) Continuous records are made of the natural potential differences between the ends of two straight insulated wires, from 1 to 10 km. long, orientated north-south and east-west. Each wire ends in two electrodes buried in the ground, and the potential differences between the electrodes are recorded by suitable voltmeters (potentiometers). In this way the measuring system does not carry a current, and interferes as little as possible with the natural conditions. (*b*) In addition, the distribution of the conductivity of the soil is determined by special surveys. The earth's resistivity below the uppermost layers depends only on geological conditions, and therefore varies only slowly with time; but changing meteorological conditions, especially

Ee

moisture and freezing, affect the electric properties of the top soil down to a few metres depth, and are reflected in the records along the lines. In addition to these 'short-line' records, from electrodes a few km. apart, earth-currents are also recorded on long telegraph lines of 100 km. length and more.

13.2. Earth-resistivity, homogeneous soil. In the case of stationary currents in a homogeneous conducting medium of specific resistivity ρ, the potential gradient (expressed in volt/cm.) is ρ multiplied by the current-density (amp./cm.²). Therefore the unit for ρ is (volt/amp.) × cm. = ohm × cm. (Sometimes ρ is defined as the resistance per cubic centimetre of the material, meaning, of course, that a cm. cube, with plane electrodes of ρ volt potential difference placed on opposite faces, transmits a current of 1 ampere.) The *conductivity* is the reciprocal of the resistivity. In the c.g.s. system of electromagnetic units, in which Maxwell's second law is written $4\pi E/\rho = \text{curl } H$, resistance has the physical dimensions of a velocity (cm./sec.), and resistivity therefore of cm.²/sec.; 1 ohm cm. = 10^9 cm.²/sec. In electrostatic units (used in atmospheric-electric work),

$$1 \text{ ohm cm.} = 1 \cdot 11 \times 10^{-12} \text{ sec.}$$

The simplest scheme of measurement is due in principle to Wenner [11]. At four points, uniformly spaced at successive distances a along a straight line, electrodes are placed in good contact with the soil (Fig. 1). A measured current of I amperes is sent through the outer pair of electrodes, A and D, and the potential difference V volts is determined between the two inner electrodes B and C. Then

$$\rho = 2\pi a V/I, \tag{1}$$

provided the depth d and the dimensions of the electrodes are small compared with a. To prove this statement we consider the distribution of the currents between A and D. This will be approximately the same as if all space was filled with matter of resistivity ρ, and an equal current I entered an electrode at A′, and left at D′, where A′ and D′ are the image points of A and D relative to the earth's surface (treated as an infinite plane). The currents being steady, and ρ uniform, the current-intensity i and the electric intensity E will be given by

$$\rho i = E = -\text{grad}\,\Phi, \tag{2}$$

where at any point P the electric potential Φ is given by

$$\Phi = \frac{\rho I}{4\pi}\left(\frac{1}{\text{AP}} + \frac{1}{\text{A'P}} - \frac{1}{\text{DP}} - \frac{1}{\text{D'P}}\right). \tag{3}$$

This is because Φ is the sum of contributions from each of the four current 'sources' (A, A') and 'sinks'·(D, D'), as if each was alone in the field and the current passed radially outwards from it to infinity, or inwards to it from infinity; for such an isolated source the current-intensity i at a distance r must be such that the total flow over the spherical area $(4\pi r^2)$ of this radius is I, so that $i = I/4\pi r^2$; hence $-\rho i$
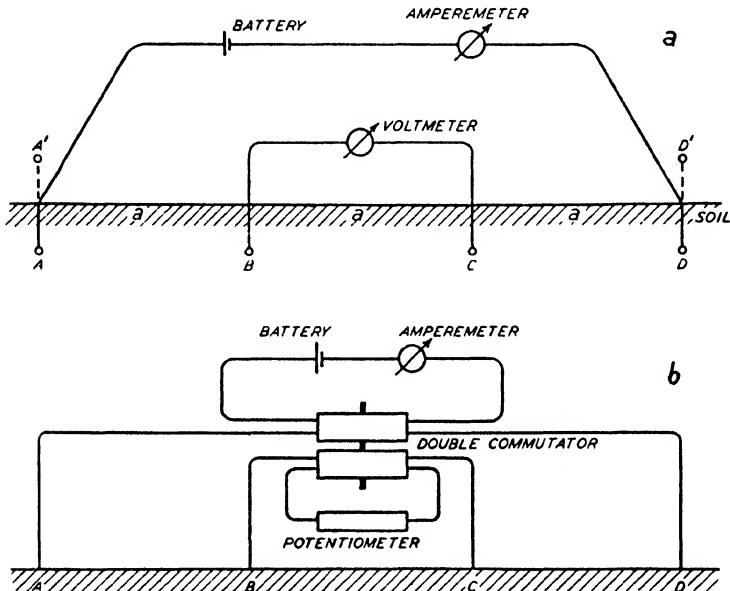
FIG. 1. (a) The Wenner method of measuring earth-resistivity, (b) The Gish-Rooney apparatus

is the gradient of the potential $\rho I/4\pi r$, or of $\rho I/4\pi \mathrm{AP}$ at a point P, if the source is at A.

Thus the potential difference between B and C is given by

$$\Phi_B - \Phi_C = \frac{\rho I}{4\pi}\left(\frac{1}{\mathrm{AB}} + \frac{1}{\mathrm{A'B}} - \frac{1}{\mathrm{DB}} - \frac{1}{\mathrm{D'B}} - \frac{1}{\mathrm{AC}} - \frac{1}{\mathrm{A'C}} + \frac{1}{\mathrm{DC}} + \frac{1}{\mathrm{D'C}}\right);$$

since d/a is small, we may take $\mathrm{AB} = \mathrm{A'B} = \mathrm{DC} = \mathrm{D'C} = a$, $\mathrm{AC} = \mathrm{A'C} = \mathrm{DB} = \mathrm{D'B} = 2a$, which gives

$$V = \Phi_B - \Phi_C = \frac{\rho I}{2\pi a}, \tag{4}$$

i.e. equation (1).

13.3. Inhomogeneous soil. When I and V are measured, the value of ρ given by (1) may be called the 'equivalent resistivity'. This would equal the actual resistivity if the soil were homogeneous;

otherwise it is a certain average $\bar{\rho}$ to which the actual resistivity of each unit volume of the soil contributes, with a weight depending on the geometrical distribution of resistivity and also on the spacing a of the electrodes. It is clear that, in general, with increasing distance a of the electrodes, the influence of the electrical properties at greater depths increases also, so that $\bar{\rho}$, plotted as a function of a, may be interpreted as indicating whether ρ changes with increasing depth z.

Little or no current, however, would penetrate below a horizontal layer that is either highly conducting or highly insulating compared with the other soil. Suppose, for instance, that a highly conducting layer of uniform resistivity ρ' extends from the surface to the depth d, and that the material below this level is completely insulating. For distances a greater than a few multiples of the depth d, the currents would spread between the electrodes A and D, as in a conducting plane instead of a semi-infinite space. A simple calculation similar to the one given in § 2 shows that then, for any given current I, the potential difference V between the points B and C would be independent of a, namely,

$$V = \rho'dI\log_e(2/\pi), \qquad \rho' = V/dI\log_e(2/\pi). \tag{5}$$

If in this case ρ was calculated by substituting the measured values of V and I in (1) it would seem that ρ increased proportionally to a, for large values of a. If this result was correctly interpreted as indicating a layer of uniform density ρ' down to some depth d, with insulating matter below, the product $\rho'd$ could be calculated from (5). Since ρ' could be found separately, by using distances a considerably smaller than d, the depth d could also be determined.

The distribution of the potential over the surface, due to a single current-pole (source) A, assuming horizontal stratification of the soil (that is, the resistivity ρ is a function $\rho(z)$ of the depth z only), is important for the theoretical interpretation of earth-resistivity measurements, because the case of two or more poles can be deduced from it. The observed surface potential-gradient $\partial\Phi/\partial r$ at the distance r_A from A, for a *homogeneous* soil in which $\rho = \bar{\rho} = $ const., will correspond to that of a source at A and an equal source at the image point A', as in § 2, or, since A' is nearly coincident with A, it will correspond to a source of current $2I$ at A. Consequently

$$\partial\Phi/\partial r_A = -2\bar{\rho}I/4\pi r_A^2,$$

or
$$\bar{\rho} = -(2\pi r_A^2/I)\partial\Phi/\partial r_A. \tag{6}$$

If actually there is stratification, so that $\rho = \rho(z)$, and if $\bar{\rho}$ is calculated from measurements of I and $\partial\phi/\partial r_A$ by using (6), the values found would

vary with r_A. The geophysical problem is the connexion between $\bar{\rho}(r_A)$ and $\rho(z)$.

The simplest way is to assume some function $\rho(z)$, work out the corresponding potential distribution, and the corresponding function $\bar{\rho}(r_A)$. This has been done for the case of up to four horizontal layers of different resistivity, by straightforward but laborious application of the method of electrical images, for instance, by Hummel [12] and Roman [13]. General forms of $\rho(z)$ have been studied by Muskat [14] and King [15].

The inverse problem, which theoretically better meets the requirements of practical geophysics, and is of general mathematical interest, is to find $\rho(z)$ from $\bar{\rho}(r_A)$; it has been solved by Slichter [16] and Langer, who expressed $\bar{\rho}(r_A)$ and $\rho(z)$ as power series, and by Stevenson [17], in terms of Fourier series.

While the case of horizontal stratification leads to determinate solutions of the inverse problem mentioned, the general case of inhomogeneous distribution of ρ, varying both with the depth and in horizontal planes (as in the case of metallic ore-bodies in rocks), has no unique solution, and the interpretation of the results of the electrical survey must depend on additional considerations, such as a knowledge of the possible geological structures involved may suggest.

13.4. The apparatus of Gish and Rooney. These authors [18] developed the following procedure for the Wenner method (Fig. 1 *b*): Radio batteries provide the current, with a total electromotive force of about 150 volts. The electrodes are steel rods, about half an inch in diameter and 2 feet long; they are driven to about 1 foot depth into the soil; where bare rock-surfaces prohibit their use, flat coils of copper wire, embedded in cotton soaked with copper sulphate, are used instead. The current I between the electrodes A and D is measured by a direct-current milliammeter having five ranges, between 0 to 50 and 0 to 2,000 milliamperes. The voltage V between the electrodes B and C is measured by a direct-current potentiometer indicator with ranges up to 1,000 millivolts. When the interval a between the electrodes is large, the resistance at the electrodes is reduced by inserting a number of electrodes near A and D. A double commutator is inserted into the lines between the meters and the electrodes; the current sent through the outer electrodes is reversed about 60 times per second by one section of the commutator while the other section rectifies the potential difference between the intermediate electrodes at the same time, so that the voltage can be read by a direct-current potentiometer. The

commutator minimizes the effects of polarization at the electrodes, and eliminates the effect of natural or artificial stray currents in the earth.

As a numerical example of the quantities involved in equation (1), suppose that $a = 500$ m. $= 5 \times 10^4$ cm., $I = 1$ amp., $V = 0.1$ volt, $\rho = 3 \times 10^4$ ohm cm. With an electrode separation of 300 m., values of ρ from 4×10^3 to 2×10^4 ohm cm. can be measured with errors under 1 per cent.; these limits of ρ vary in the same sense as the electrode separation.

A commercial apparatus sometimes known as a 'Megger' (a name derived from megohm $= 10^6$ ohm), which is used for testing insulation resistance in power or telephone lines, is so constructed that the coils of the current-meter and the voltmeter are connected so as to give direct, though only approximate, readings for the ratio V/I, so that ρ can be easily deduced. Experimental uncertainties arise from the imperfect working of the commutator; manual reversal of the current and potentiometer every few seconds has been recommended.

13.5. Other methods. Many other methods have also been used in geophysical prospecting. A slight alteration of the method described is the 'five-electrode method', in which a fifth electrode at the centre of the line is inserted, from which the potential differences are measured. Schlumberger and others use current led into the ground through two electrodes separated by a distance comparable with the supposed depth of the ore, and determine the lines of equal potential on the surface by means of two movable search-electrodes, say, E and F, connected by a cable in which a micro-voltmeter near F is inserted. In order to avoid contact potentials or polarization effects, these search-electrodes are of the 'non-polarizing' type, for instance, a copper rod in a concentrated solution of copper sulphate contained in a porous pot of the type used in primary cells, or a silver foil bent to the cylindrical form in a solution of sodium chloride [20]. One of the search-electrodes, say E, is placed in the ground, while the observer moves the other electrode F until the micro-voltmeter reads zero. Then E and F lie on the same equipotential line. This procedure is continued until a sufficient number of such equipotential lines are plotted, the theoretically assured closure of the lines forming a good test of the accuracy of the observations. Lundberg [21, 22] feeds the current through parallel linear electrodes. Models of the assumed geological structure in laboratory scale are tested and changed till they yield the same shape of the equipotential lines as found in the field.

The problem of power-lines with currents returning through the earth,

thereby interfering with telephonic communication lines, has directed the attention of engineers to the distribution of currents set up in the earth by alternating electromotive forces [9 and 20]. This distribution depends largely on inductive effects, similar to the skin effect in metallic conductors. The essential coupling between two parallel lines is expressed in terms of a coefficient of 'mutual impedance'; experiments show that values for the resistivity of the earth inferred from actual measurements of this coefficient agree, up to frequencies of 3,000 cycles/sec., with those found from direct-current measurements. For wireless frequencies above 60 kilocycles/sec., the theory has also to take into account the 'displacement currents' in the earth, which can be neglected for lower frequencies.

Alternating currents, induced in the earth by currents in overground lines and detected by search coils or frame aerials, have also been extensively used in geophysical prospecting for metal ores, oil, or water. By perfection of the operating technique Slichter [23] has made surveys over distances of more than 100 miles. For detailed descriptions books on applied geophysics [24–31, 4.1, 2.4] should be consulted.

13.6. The general results of earth-resistivity surveys. The earth-resistivity ρ as determined by the methods described may vary between 10^2 and 10^8 or more ohm cm. The values for the resistivity of minerals or small samples of rocks measured in the laboratory are, in general, not directly comparable with the values measured in the natural structure of soils and rocks, because the water-content, pressure, and structure of the samples are not representative. The following average values give the order of magnitude of ρ, together with the values for some pure substances, inserted in brackets; for references, see the papers by Reich [27].

TABLE 1.   *Specific resistivity in ohm cm.*

(Pure copper	2×10^{-6})
Magnetite	1
(Concentrated NaCl solution	5)
Sea-water	21
Moist loam	100 to 5,000
Top soil, average value used in electrotechnics . .	10,000
Moist sand	10^5 to 10^6
Chalkstone, sandstone	10^5 to 10^6
Dry sand, granite	up to 4×10^6
Rock salt in natural deposits	10^7
(Pure rock salt	10^{17})
Atmosphere near the ground	6×10^{15}

Because of the importance of sea-water for induction problems some values of ρ are given for various temperatures and salinities (after

Krümmel [32]); the salinity is defined, as is usual in oceanography, as the total amount of solid material in one kg. of seawater after all the carbonates have been converted into oxides, the bromine and iodine replaced by chlorine, and all organic matter completely oxidized. The salinity 35 grams per kg. is typical for the open ocean.

TABLE 2. *The resistivity of sea-water in ohm cm.*

Tempera-ture (° C.)	Salinity					
	5	*10*	*20*	*30*	*35*	*40*
0	209	109	57	39	34	30
10	169	82	43	30	26	23
20	126	65	34	24	21	18
30	103	53	28	20	17	15

Extremely pure fresh water, protected from the air, has been found to give resistivities as high as 2×10^7; even small amounts of electrolytes lower this value considerably, for instance, to 4,000 ohm cm. in the Potomac River near Washington, D.C. [18], a value which is also fairly representative for ordinary ground-water. Salt water (for instance, ground-water in desert regions) lowers ρ to the values given in the table, and makes such water horizons especially accessible to geophysical prospecting. Changes in the water-content of porous soil, and changes in the electrolytic content of this water affect ρ considerably; in fact, the resistivity of moist strata is much nearer to the low value for ground-water than to the high value for dry rock.

For prospecting, resistivity-measurements have certainly been used with advantage to locate certain underground discontinuities, for instance, underground water-level, or the depth of overburden over minerals or ores. Older reports of investigations connected with commercial prospecting firms should, however, be read with a certain reserve. Other electrical methods of prospecting are described in the text-books cited [24–37].

13.7. Examples of resistivity surveys. (*a*) Gish and Rooney [18] tested their apparatus on the Potomac River over a depth of 9 to 10 feet. The average results, shown in Fig. 2, show for small electrode-separations the average uniform resistivity 4,400 ohm cm. of the water, as confirmed by laboratory measurements: then (after the electrode-separation became greater than the depth) a decrease of the equivalent resistivity ρ was observed, because of the higher conductivity of the mud or silt at the bottom; and finally there was a rise of the equivalent ρ to the higher value for the underlying material.

(b) In grading operations near Washington, D.C., a ravine of about 500 feet length and 30 feet depth had been filled with material cut from the hill at the side. Resistivity-measurements on a line across the ravine were made with electrode-separations of 25 to 300 feet; the values of ρ thus found are shown in Fig. 3 at the position of the centre of each line measured [18]. The higher resistivity (up to 80,000) of the loose material is clearly indicated, especially for the smallest electrode-separation $a = 25$ feet, and the curves fit well with the profile of the
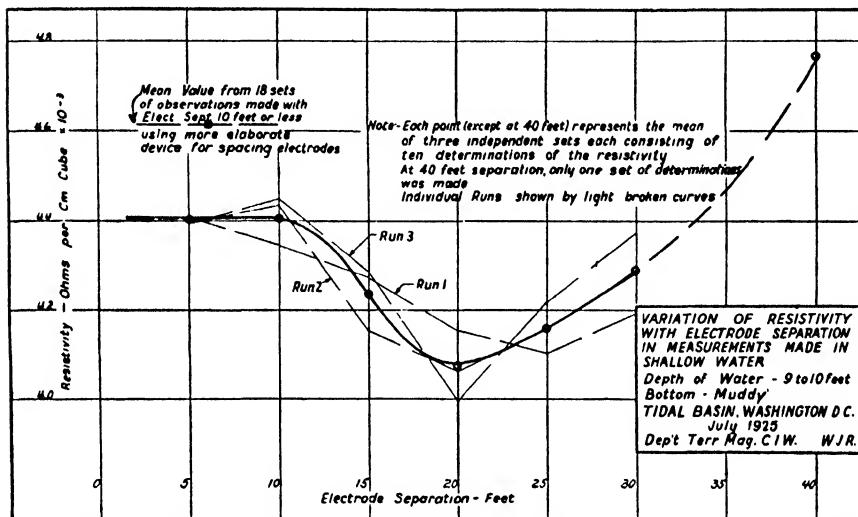
FIG. 2. The variation of electric resistivity (ordinates) with electrode-separation (abscissae) in measurements made in shallow water (Potomac river; depth of water 9 to 10 feet). (After Gish and Rooney)

fill. When a is increased beyond 100 feet, ρ rises again, probably indicating the still higher resistivity of the bed-rock.

(c) In both cases the critical electrode-separation a, at which a change of ρ with increasing a was found, was approximately equal to the depth of the sub-surface discontinuity. In measurements by Stern [33], made in the lignite region of the 'Ville' in the lower Rhine valley, the equivalent resistivity was measured in arbitrary units, and plotted against the electrode-separation a, which was also interpreted as the depth. The discontinuity between the overlying gravel and the lignite is, in this particular case, indicated by a maximum of the equivalent resistivity. These simple inferences from field-work have, however, been seriously doubted and have certainly not yet been satisfactorily brought into line with theoretical calculations of potential distributions [15].
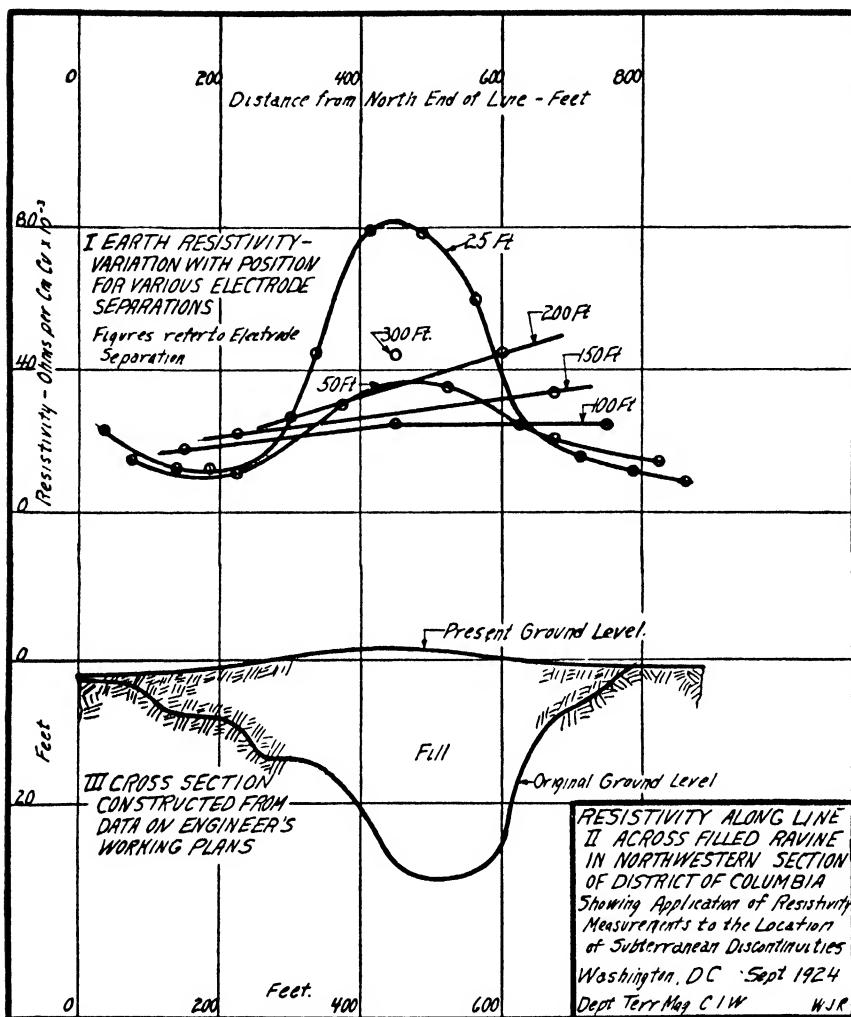
Fig. 3. The electric resistivity along a filled ravine (cross-section with 10-fold vertical magnification shown below) for electrode separations 25, 50, 100, 150, 200 and 300 feet. (After Gish and Rooney)

(*d*) A decrease of resistivity of chalk-stone during three rainy weeks has been traced by Löhnberg and Stern [34].

(*e*) The variation of resistivity with depth (or electrode-separation) has been measured by Rooney [35] for observatories where continuous earth-current records are made. At Ebro, Spain, ρ increases from the low surface values in the river valley (of the order of 2,000 ohm cm.) to values of about 12,000 ohm cm. between 100 and 300 m. depths. At Watheroo, Western Australia, ρ for the dry sand at the surface has

values up to 4×10^6 ohm cm., but drops to about 700 ohm cm. for $a = 100$ m., increasing slowly to 2,000 ohm cm. for $a = 300$ m. and 5,000 ohm cm. for $a = 600$ m.

At Huancayo, Peru, the mean values of ρ to depths of 200 or 300 metres are about 10,000 ohm cm., which is typical for underlying sedimentary rocks.

Rainfall after a dry season decreases the surface resistivity sharply, but for depths greater than about 10 metres no appreciable seasonal or rainfall effect can be detected.

(f) By assigning to each geological formation its average resistivity, as found in a sufficient number of field tests, geological maps can be converted into maps of resistivity. The Electrical Research Association, with the assistance of Broughton Edge, has published such a chart for Great Britain, showing the equivalent resistivity corresponding to about 500 feet in depth, subdivided into five ranges with the limits 1,000, 3,000, 30,000, and 300,000 ohm cm. [20]. Slichter [23] has prepared such an electrical map for Central Massachusetts.

13.8. Earth-current observatories. The simplest means of observing earth-currents continuously would appear to be by connecting two grounded electrodes, for instance, water-pipes, by an insulated wire, and to record the currents in the wire by inserting a registering galvanometer. The first circuits specially built for this purpose were established at Greenwich Observatory in 1865, and consisted of two lines approximately 15 km. long, extending north-south and east-west. These and several other European lines installed in the second half of the nineteenth century had to be discontinued later, because of the disturbing effects of stray currents due to electric railways and other sources. After 1900 it seems that only the Ebro Observatory, Tortosa, Spain, maintained earth-current records (for two lines of about 1·4 km. length), until Bauer decided to instal earth-current measuring systems at the two observatories of the Carnegie Institution of Washington at Watheroo (Western Australia) and at Huancayo (Peru), far from all interference by power lines. These systems, planned and installed by Gish and Rooney [38, 39], eliminated some serious errors which affected former observations. They are described in § 9, and the subsequent discussion of earth-currents will refer mainly to their records, in preference to the more qualitative older observations, because, as Gish pointed out, 'in the light of our present knowledge of correlated branches of cosmical physics, earth-current phenomena should not present such baffling complexities, nor with the improved instrumental

equipment now available should the difficulties of reliable measurement prove as discouraging, as in the years when interest in this subject was at a maximum'. Similar systems were installed in 1931 at Tucson, Arizona [40] and, for the International Polar Year of 1932–3, at College-Fairbanks, Alaska [41].

13.9. The earth-current measuring system. An earth-current observatory should represent the conditions over a considerable area of the globe, and should therefore not be placed in a situation which is affected by local peculiarities. Hence an earth-resistivity survey should first be made in order to choose a site at which the resistivity is laterally isotropic or, in other words, such that the geological structure does not force the earth-currents to flow preferably along certain lines of low resistivity. The lines should be approximately level and at right angles to each other, and should preferably be in the north-south and east-west directions.

Spurious effects which cause potential differences between the electrodes and the earth are the main difficulty in earth-current measurements. Metallic electrodes of exactly the same material and form, placed in contact with the soil at adjacent points, usually show appreciable differences of potential (between a few hundredths and a few tenths of a volt). Mauchly [42] filled a glass tube, 150 cm. long and 5 cm. wide, with freshly dried, sandy soil, and embedded two zinc electrodes near the ends. There was an average potential difference between them, of about 40 millivolts, one end (say A) being positive relative to B; but when the tube was held vertical, with A below, the potential difference was 53 millivolts, whereas with A above it was only 26 millivolts. Other experiments confirmed the fact that, besides a non-reversible electromotive force, there was a reversible force of 10 to 15 millivolts, making the lower end positive; this was ascribed to the effect of the pressure difference. Temperature effects of the same order were found.

Part of the electromotive force arising at a metallic electrode put into soil is caused by a thin film at the surface of the electrode, and part arises from changes in the solutions surrounding the electrode. Among a number of metals compared, lead, iron, and possibly cadmium were found to be the best suited for electrodes. Non-polarizing electrodes such as were mentioned in § 5 are not suitable if long records over many years are contemplated, because the leaching of some solution from the porous pot into the soil may gradually lower the resistivity in the vicinity.

No perfectly permanent type of electrode has been found, but the following type, used at Huancayo and Watheroo, has proved satisfactory. Each electrode is a horizontal web-shaped grid, about 6 metres in diameter, built up of five concentric circular loops of lead wire connected by eight radial arms which meet at the centre to form a core, where the connexions from the line enter. Each electrode was placed at a depth of 1·5 metres or more, where the daily temperature-variation is less than 0·1° C. Two electrodes were placed side by side at each terminal, in order to study electrode effects. Care was taken that all electrodes were placed into the same type of soil, and that their contact resistances are nearly equal.

The two lines form a right angle or a cross, and run exactly north-south and east-west. Their lengths are between 1·6 and 3·4 km. Although spurious electrode effects would be less troublesome with lines of 100 km. length or more, these shorter lines were preferred because they are more easily supervised, and because the geology of the underground is likely to be more uniform in smaller areas. The eastward line at Watheroo is longer than the northward line, in order to magnify the (smaller) east component of the gradient. Two separate crosses are laid out at Huancayo. Each component is measured at least on two separate lines, and this duplication ensures continuity of the records even if a line breaks down occasionally.

At Watheroo both overhead and underground lines were constructed for comparison. No differences were apparent in the results except during thunderstorms, when typical disturbances are found in the records from overhead lines, while the underground lines are entirely unaffected.

The recording apparatus is a special type of Leeds and Northrup multiple-point recorder, employing the compensation method for measuring the potential; the electromotive force is balanced, by a slide-wire arrangement, against the standard cell. It is operated by a constant-speed motor. Between six and twelve circuits can be switched on in succession. The sensitivity is adjustable so that the recording sheet, 25 cm. wide, accommodates a change of potential of either 80, 800, or 8,000 millivolts. A point on the record is made about every 50 seconds, so that, with six different lines, voltage on the same line is recorded every three minutes. Hourly means are formed, just as for terrestrial magnetic components.

For the College-Fairbanks (Alaska) station Gish [43] has devised a somewhat different recording apparatus. Each line is connected with

a photographically recording d'Arsonval moving-coil-type galvanometer with current-sensitivity of 10^{-10} ampere per mm. deflexion. The coil resistance is 500 ohms; the galvanometer is shunted with the 'critical damping resistance' of about 21,000 ohms, which damps the free oscillations of the coil so that it approaches its position, corresponding to the current, asymptotically. This damping resistance is provided with taps in order to make the sensitivity adjustable. A constant high resistance—2 million ohms (usually termed 2 megohms) —is inserted in the series in order to throttle the current actually flowing in the circuit to the order of 10^{-8} ampere. This series resistance is chosen so as to be high relative to the resistance of the earth between the two electrodes, so that the current lines in the earth are not distorted, and also to be high relative to the contact electrode resistances (between 10 and 10,000 ohms), so that changes at the electrodes do not affect the recorded values. The calibration to electromotive forces is made by switching over the galvanometer, from the line connecting the field electrodes, to a potentiometer arrangement which applies a known electromotive force.

13.10. The earth-current components. According to an international convention [44], a flow of earth-current along the (astronomical) meridian is designated as the *northward component N*, and is reckoned positive when the flow is towards the north, and negative when in the opposite sense; similarly the other rectangular component is the *eastward component E*. The potential differences are expressed in millivolts per kilometre. The two rectangular components can be combined vectorially, as usual, to give the resultant flow, which is of magnitude $R = (N^2+E^2)^{\frac{1}{2}}$, and of azimuth α (reckoned eastwards from the north), where $\tan\alpha = E/N$. These relations are derived on the assumption that, at any moment, the potential gradient is uniform over the area covered by the measuring circuit. If this is so, then the successive equipotential lines drawn for unit differences of potential are parallel and equidistant (Fig. 4 a), and R is equal to the number of lines per unit length along the perpendicular (OR) to the equipotential lines; for a straight line drawn with the azimuth β,

$R\cos(\alpha-\beta)$ = potential difference per unit length along azimuth β. (7)

Putting $\beta = 0°$ and $90°$, we obtain, as above,

$$N = R\cos\alpha, \qquad E = R\sin\alpha. \tag{8}$$

If for local reasons the directions ON', OE' of the measuring circuits cannot be made exactly north or east, but form angles δ and ϵ with the

true directions ON, OE (Fig. 4 b), the potential differences per unit length, say N' and E', can be converted into N and E as follows, since

$$N' = R\cos(\alpha-\delta), \quad E' = R\sin(\alpha-\epsilon),$$
$$N = (N'\cos\epsilon - E'\sin\delta)/\cos(\delta-\epsilon),$$
$$E = (N'\sin\epsilon + E'\cos\delta)/\cos(\delta-\epsilon). \tag{9}$$

13.11. The average earth-current. Usually the daily averages of the absolute potential differences measured between two electrodes have no general geophysical significance; they depend merely on the local conditions at the electrodes. Rooney [45] cites the following
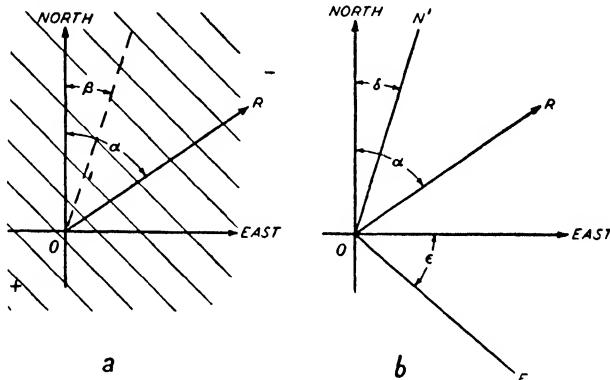
FIG. 4. Illustrating components of electric earth-currents (in a, R should be perpendicular to the parallel lines)

extreme example: In March 1926 the potential between the common electrode O at Watheroo and an electrode P', one mile north of O, was -28 millivolts, while the potential between O and P'', another electrode less than 60 feet from P', was $+41$ millivolts; these potentials, erroneously interpreted as due to actual earth-currents, would indicate a large flow from P' to O in the first case, and an even larger flow from O to P'' in the second case, which is, of course, absurd.

In Watheroo the mean daily potentials vary with the meteorological conditions of the wet and dry seasons. They are entirely ascribed to the contact-potentials at the electrodes, which change with the amount and salt content of the moisture in the soil; therefore the mean potentials between any two electrodes are independent of their distance apart. An actually constant earth-current (constant throughout a day) could therefore be detected only if the distance between the electrodes could be made so great that this average natural current would be certainly greater than the contact-potentials. So far, for reasons of economy, only the long lines of telephone or telegraphic services have

been available for such measurements, and since their construction is not suitable for exact tests, no satisfactory results have been obtained; that is to say, the measurements do not give conclusive evidence of a constant part in the earth-currents, which is greater than the contact-potentials of the electrodes on long lines, that is, greater than about 0·1 volt per 100 km., or 1 millivolt/km. In the discussion of earth-currents, therefore, we have so far no counterpart to the planetary permanent magnetic field of the earth; our discussion is confined to the variations occurring within such time limits as to exclude changes
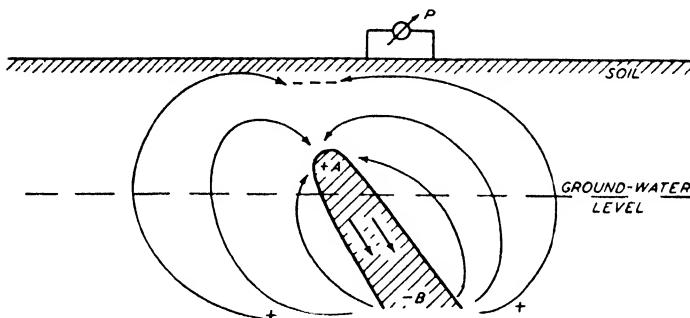
Fig. 5. Illustrating electric earth-currents due to spontaneous polarization

of the contact-potentials. These variations are the regular daily changes, and the irregular changes during magnetic and earth-current storms.

The first indication of a real change in earth-currents, with a period longer than one day, is afforded by Rooney's [46] discovery that the daily mean of the potential gradient for a very disturbed day differs from that for a markedly calm day by about 0·1 millivolt/km. at Tucson (Arizona); this amounts to a few per cent. of the amplitude of the daily variation; at Huancayo an effect of similar relative magnitude is found. The additional average current on disturbed days is directed toward the north-east at Tucson, and to the south-east at Huancayo; these directions may be modified by local geological features.

13.12. Spontaneous polarization. An exception to the statement in § 11, that constant earth currents or potential differences cannot be observed, is found in the small-scale 'self-potentials' measured over certain ores. Consider (Fig. 5) a copper sulphide ore under an overburden of 20 feet. Owing to the moisture of the soil the upper portion of the vein is oxidized. Thereby an electrochemical potential difference of half a volt or more arises between the upper and lower portions, and currents flow downward in the well-conducting vein, and upward in

the poorly conducting surrounding soil, as in a battery. The current-lines near the surface converge to a 'negative centre' above the ore, and the equipotential lines surround this centre. These potential differences can be found by means of non-polarizable electrodes, much in the same way as in the case of an artificial current-supply (§ 5). The method is restricted to sulphide ore-bodies of a pyritic character, and fails even there if an impervious clay overburden prevents percolation and, thereby, active oxidation. But in many circumstances spontaneous polarization is a good indicator; as early as 1830 Fox [47] placed large copper plates as electrodes against the rock inside Cornish mines, measured the natural current with a galvanometer, and located large deposits.

Much smaller in individual size and intensity, though perhaps a more common phenomenon, are the 'turbulent currents' measured by Hunkel [28] with non-polarizing electrodes; they appear as irregularly distributed positive and negative poles of a few millivolts, at distances of the order of a foot. They seem to be connected with small inequalities in the chemical disintegration of the soil, causing local differences in the 'acidity' of the soil solutions. Near faults, the disintegration extends deeper, and more intense poles can be expected.

13.13. The problem of vertical earth-currents. Between two earth-electrodes placed in the soil at different altitudes on a hill-side, there is generally an average potential difference of the order of 100 millivolts per km. length of the line, with a current flowing in the connecting wire toward the upper station. This current seems to be more constant and relatively less variable, the steeper the line; its variations do not change in accordance with the terrestrial-magnetic variations, as do the earth-current variations in horizontal lines [2, 48–50]. But observations on a line extending up the side of Vesuvius revealed that the current, which flows upward during volcanically quiet times, is reversed during times of volcanic activity. At the Ben Nevis Observatory in Scotland a strong current which flowed upward when the summit was clear, was reversed when fog or clouds enveloped the summit.

Gish [49] is sceptical as to these measurements. If the currents measured in the wires do actually flow in the soil from all sides of the mountain top, the circuit must be closed somehow. If it is assumed that the gradient has the same upward direction throughout the mountain, then the current should be closed through the air; a large current would go into the air on the summit. This, however, is exactly the opposite direction of the air-earth current observed in atmospheric electricity.

In fine weather the downward potential gradient in the air is about 1 volt/cm., which, with an average resistivity in the open air of 6×10^{15} ohm cm., results in a current-density 2×10^{-16} amp./cm.^{2} flowing from the air *into* the ground. The magnitude of the apparent 'uphill' earth-currents, with a voltage of, say, 0·2 volt per km. height difference, or 2×10^{-6} volt/cm., flowing in earth of resistivity 10^5 ohm cm., would be 2×10^{-11} amp./cm.^{2}, that is, 100,000 times more intense than the actually observed air-earth current.

This leaves only two alternatives: either we must assume that the 'uphill' currents form closed circuits with currents flowing downward in the interior of the mountain, or that they only represent systematic electrode effects due, for instance, to a variation of acidity (hydrogen-ion content) of the soil with height, depending on height-differences in erosion or weathering. In the latter case, the mountain resembles a battery in which a current flows only when the circuit is closed by the electrodes and the wire between. No final decision has been reached as to such possible electrochemical or thermo-electrical processes.

Lemström [10], as long as fifty years ago, put his 'point-apparatus' on the tops of the Finnish mountains; it consisted of barbed wire (fastened on insulators) covering several hundred square metres of the top, and was connected, through a galvanometer, with an earthed plate in the valley. The numerous point-discharges—at night sometimes visible as Saint Elmo's fire—set up a downward current, which is readily explained as a consequence of the normal atmospheric-electric vertical current. Similar but much weaker currents are observed in the wires leading to ordinary aerials (antenna-currents).

13.14. The general character and magnitude of earth-current variations. At a number of places where earth-currents have been observed, among them Watheroo, Huancayo, Ebro, Paris (Parc St. Maur), and Batavia, the direction of the resultant current-flow is practically restricted to a single azimuth or its reverse. Although this peculiarity is absent at a few stations (for instance, at Tucson, Arizona, where the direction of the resultant flow may vary considerably), it is striking that it should be found at such widely spread localities, especially where the resistivity is found to be so laterally uniform as near Watheroo and Huancayo. In 12 earth-current systems, each about 200 km. long, operated on the cables and wire-lines of the Bell System along the eastern coast of North America from Maine to Cuba [56], this pronounced constancy of direction (there, towards the north-west or south-east) is found to prevail along the eastern part, possibly even

into Cuba; but there are three stations (in Maine, in Illinois, and in the lower Mississippi valley) where the resultant current-flows are not strictly alined, but vary in direction.

Another general feature is the close relation between the earth-resistivity of the locality and the magnitude of the earth-current variations. A remarkable example was observed at Huancayo; the east component was in the earlier years recorded on two lines, and the potential gradient was found to be nearly twice as large on the line laid where the resistivity was locally about three times as high as is generally found in the region. Since the current-density equals the ratio of the potential gradient to the resistivity, this would suggest that the magnitude of the current-density varies less, from place to place, than that of the potential gradient. In Watheroo the potential gradient on quiet days, averaged for a certain hour, is of the order 1 milli-volt/km. $= 10^{-8}$ volt/cm. which, with ρ approximately equal to 1,000 ohm cm. up to 300 m. depth, indicates a current-density of roughly 10^{-11} amp./cm.^{2} At Huancayo, the corresponding value is about 3×10^{-8} (volt/cm.)/10^4 ohm cm. $= 3 \times 10^{-12}$ amp./cm.^{2} During magnetic storms instantaneous values of the current-density may be 10 or even 100 times as high.

13.15. Earth-current storms. The simultaneous occurrence of magnetic and earth-current disturbances is the best established fact about earth-currents. Statistical discussions of the frequency of earth-current storms therefore indicate the same features as are found for magnetic storms, namely, a variation in parallel with the 11-year sun-spot cycle, a higher frequency at the equinoxes, frequency of sudden commencements after quiet intervals, and a 27-day recurrence [56] tendency.

As to quantitative relations between the two vectors, the earth-current density and the disturbance of the magnetic force, the evidence is much less conclusive. Van Bemmelen [57] compared simultaneous oscillations, having periods between 1 and 20 minutes of time, in the east earth-current component E and the north magnetic component X. He found that the oscillations have approximately opposite phases—though the highest current towards the east occurs somewhat earlier than the smallest north magnetic force—and that the ratio of the amplitudes of E and X decreases when the duration (or period) of the oscillation increases; but the ratio decreases much less rapidly than the reciprocal of the duration. However, even these results cannot be regarded as significant, because the oscillations in the north earth-
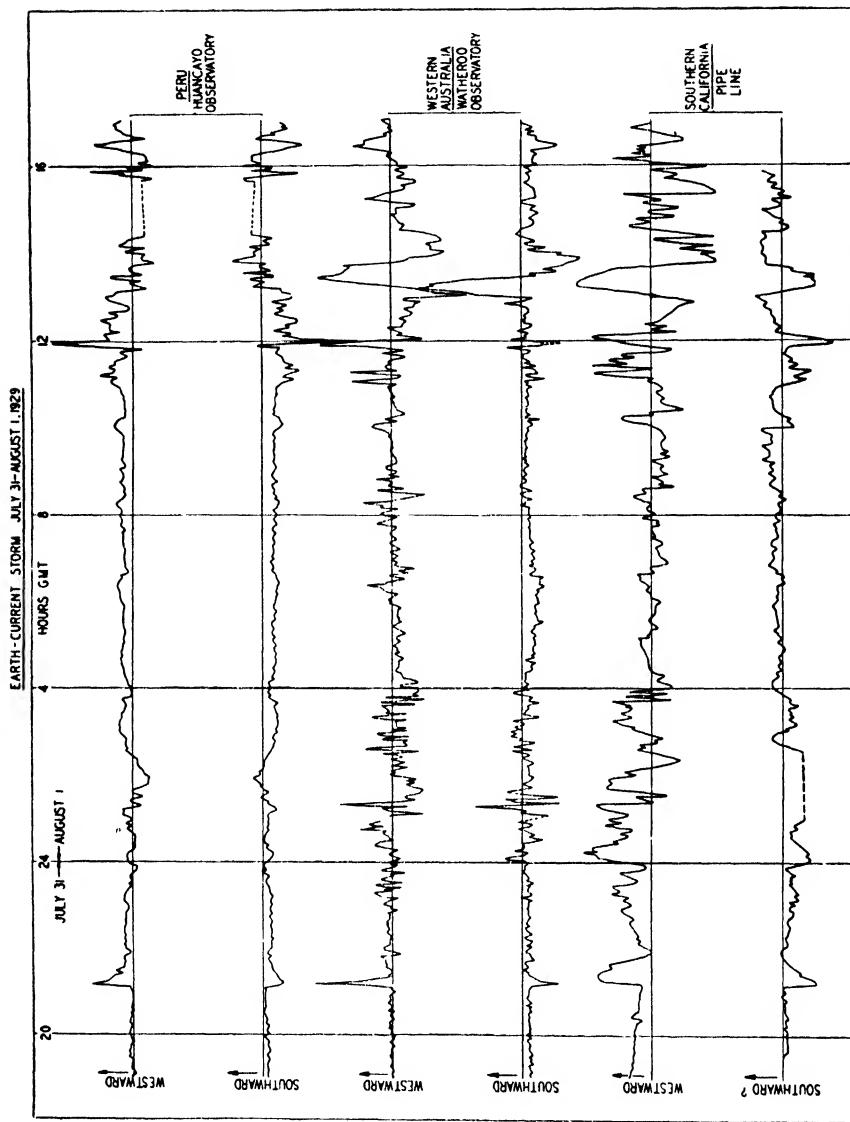
Fig. 6. Three sets of records of an earth-current storm. (After Gish)

current component N showed much closer relation to those in E than to those in the magnetic east component Y; the 'single-azimuth' character of the earth-current, described in § 14, blurs the relationship with the variations of the horizontal magnetic components which, generally, change much more in direction (9.17 and 9.18).

Nevertheless, it seems to be established that fluctuations of short period, or high frequency, are comparatively more pronounced in earth-current records (Fig. 6) than in magnetic records.

Bosler [58], from the Parc St. Maur (Paris) curves, found a parallelism between X and E, and also between Y and N; an observer looking in the direction of the earth-current would describe this relation as a deflexion of the magnetic force towards the left. Bock and Moench [59] found similarities between the northward earth-current and the variations of declination.

At stations in low latitudes the oscillatory disturbances of short periods (from a few minutes up to an hour) are of about the same order of magnitude as the normal range of the daily variation (expressed in hourly means); only during exceptional storms do they become as much as twenty or thirty times as great. At the polar station of Fairbanks, Alaska [60], the short-period disturbances extending over half an hour or less had amplitudes exceeding the normal daily variations in ratios 50 to 150. Also the absolute resultants were sometimes as high as 1,000 millivolt/km.; this might be partly due to the supposedly high resistivity of the frozen soil. The oscillatory disturbances were practically absent during the hours of daylight. During the night their intensity varied from hour to hour in good correlation with the aurora.

13.16. Daily variations. The tabulated hourly values for the earth-current components are combined in the same way as magnetic observations, in order to obtain the average daily variations, expressed as deviations from the daily mean which, as being without physical significance (§ 11), is not considered. They share the following features with the magnetic daily variations. The variations are more rapid during the daylight hours than at night; their range is higher in summer than in winter. This is most convincingly demonstrated by the Watheroo results [38], because the resistivity at Watheroo has no marked seasonal variation, and also by the Tucson results (Fig. 7). The main variation is in the north components. Observations in the northern and southern hemispheres point to a phase-change near the equator; at midday the main currents flow towards the equator from both sides. A first schematic picture of the daily current-circuits in earth-currents based
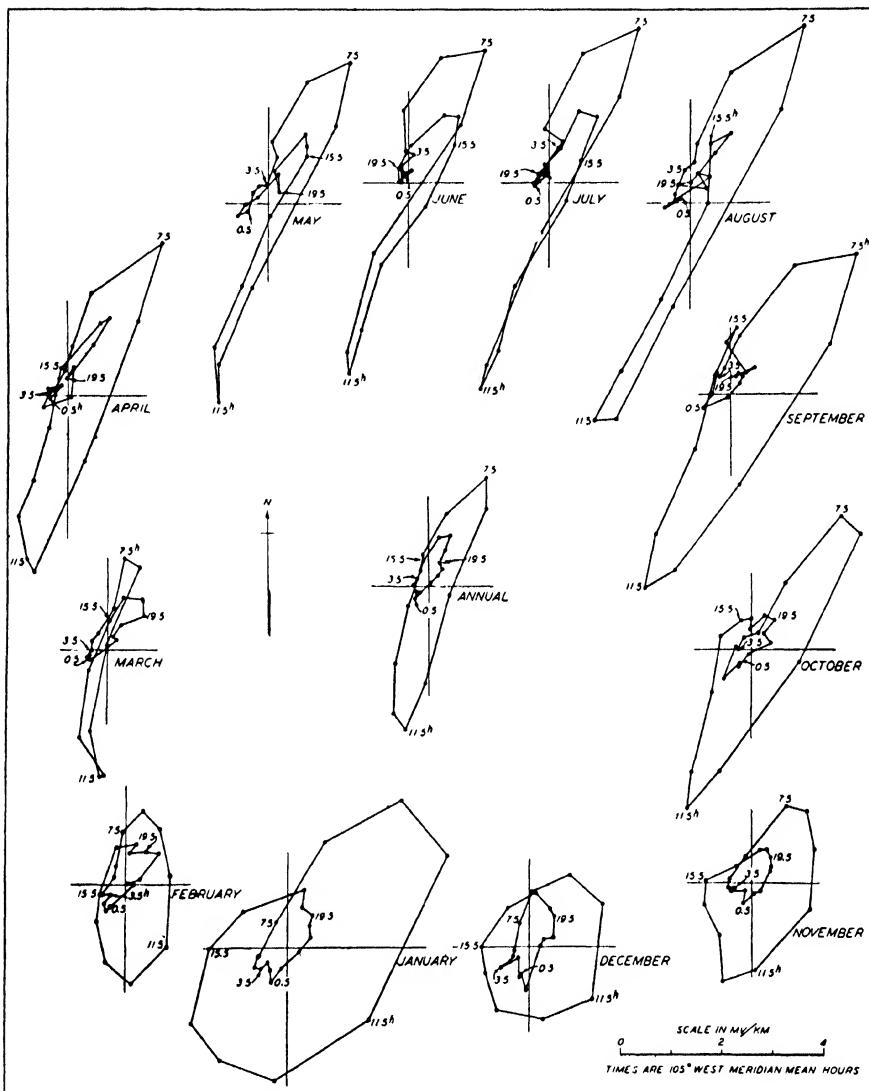
FIG. 7. The daily variation of the earth-current intensity at Tucson, Arizona, 1932-4:
monthly and annual average vector diagrams. The local times are indicated

on a collection of observational data (Fig. 8) has been derived by Gish
and Rooney [61] (Fig. 9).

Two other facts point to a physical relationship between the daily
variations of earth-currents and of terrestrial magnetism; the average
daily variations of the earth-current components, computed for the
five *magnetically* quiet and the five *magnetically* disturbed days of the

month (Fig. 10), differ by a curve which resembles a pure sine-wave
with the period of a day (lowest part of Fig. 10); this is exactly the type
of difference curve obtained for the magnetic components themselves
(see Ch. IX). Furthermore, even magnetically quiet days of the same
season may have magnetic daily variations of rather different ampli-
tudes (see 7.8). In a striking example [63] the earth-current variations
were found to be large or small on the same days (all being magnetically
quiet) as the magnetic variations (Fig. 11); note that the first and third
pairs of days follow each other with an interval of 27 days.

Recently lunar daily variations in earth-currents have been studied
by Egedal [67] for Ebro observatory and by Rougerie [68] for Paris.
Rooney [69], studying one year's records at Huancayo and Tucson,
found a well-marked lunar daily variation at Huancayo; the 12-hourly
sine-waves in the northward and eastward earth-current components
in the lunar daily variation were as large as one-sixth of the corre-
sponding amplitudes in the solar daily variation. Just as for the
magnetic lunar daily variations (Ch. VIII), those in the earth-currents
at Huancayo are magnified during the daylight hours, and are smaller by
night. This effect, however, has not been found for Tucson and Ebro.

**13.17. Physical relations between terrestrial magnetism and
earth-currents.** In many papers simultaneous observations of the
magnetic variations and of earth-currents at the same locality have been
discussed, in order to decide between the following alternatives: (a) the
magnetic variations represent the magnetic field of the earth-currents;
(b) the earth-currents are induced by the magnetic variations.

We shall consider both hypotheses on the assumption that the
observations of the earth-currents and the magnetic variations are
representative of uniform conditions over a wide area surrounding the
station. On the hypothesis (a), the magnetic field of an earth-current
of density i amp./cm.^{2}, extending to a depth d, can then be computed
as that of a homogeneous current of intensity id amp./cm. in an infinite
horizontal plane, eqn. (29), p. 19. The magnetic force above this sheet
is directed towards the right, at right angles to the current, and its
strength is $2\pi id/10$ gauss.

On the hypothesis (b), the earth-currents would be proportional to
the rate of change dF/dt of the magnetic vector F. Suppose that the
magnetic variation consists of an increase of the east component only,
that is, dY/dt is positive, while $dX/dt = dZ/dt = 0$. Suppose also that
this change is the same over a vertical rectangular area or section on
the earth, extending a km. to the north and h km. downward. If (p. 27)
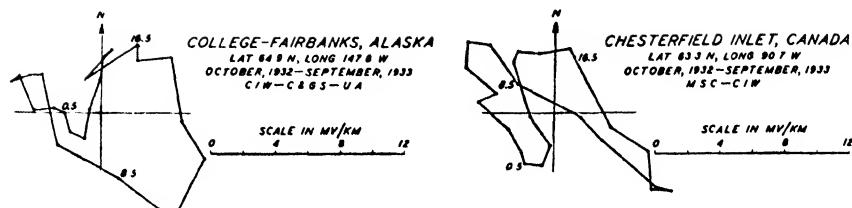
HIGH-LATITUDE STATIONS
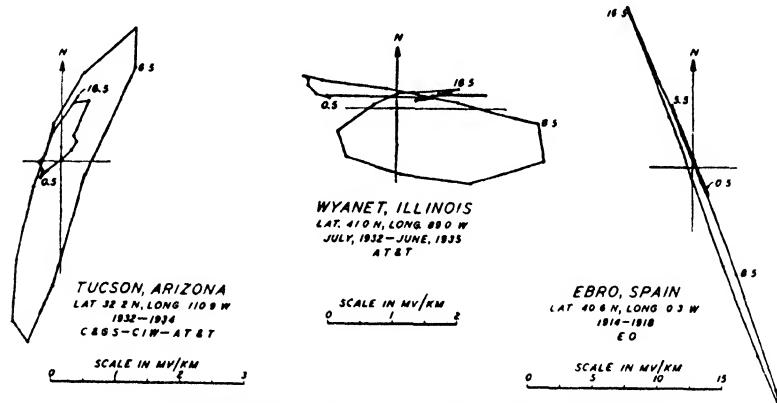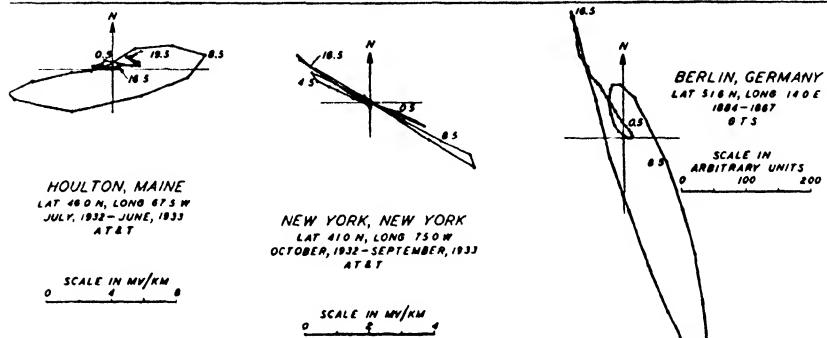
NORTHERN MIDDLE-LATITUDE STATIONS

FIG. 8 a.

Vector diagrams showing the daily variation of earth-current

dY/dt is measured in units of 10^{-5} gauss/sec. $= 1\gamma$/sec., the total voltage in a circuit around this area would be $10^{-3}ah\,dY/dt$ volts, corresponding to a gradient $10^{-3}[ah/2(a+h)]\,dY/dt$ volt/km., towards the north at the surface, and towards the south at the depth h. If h is small relative to a (say $h = 200$ km., $a = 2,000$ km.), this would give a gradient at the surface, towards the north, of $(h/2)\,dY/dt$ millivolt/km.

It is curious to note that, qualitatively, the observations of the daily magnetic variations in middle latitudes can be invoked in support of either hypothesis. As indicated in Fig. 12, the east current E varies in parallel with the south magnetic component $-X$, that is, the variations of the magnetic north component seem to be secondary induction effects of the east current E, corresponding to hypothesis (a). Quantitatively, a current-density $i = 10^{-11}$ amp./cm.² extending to a depth $d = 200$ km. would give a north component of -12γ. On the other hand, the north component N does not agree at all with the variations of the magnetic east force Y, but is closely parallel to dY/dt. Quantitatively,

$$dY/dt = 10\gamma/\text{hour} = 3 \times 10^{-3}\gamma/\text{sec.,}$$

EQUATORIAL STATION

SOUTHERN MIDDLE-LATITUDE STATION
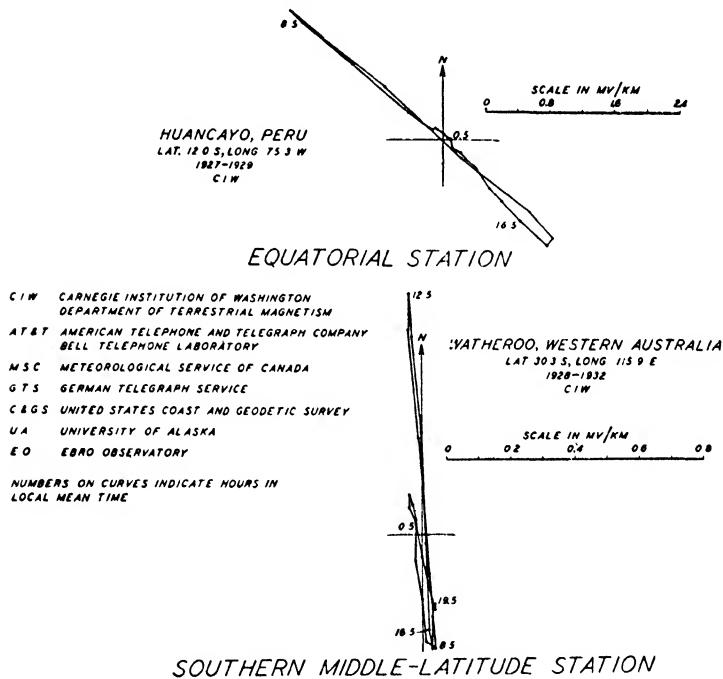
FIG. 8 b.

intensity at various stations. (After Gish and Rooney)

to a depth $h = 200$ km., would account for a potential gradient of 0·3 millivolt/km, according to hypothesis (b).

Although in both cases even the quantitative agreement is not bad, there is no question that the actual relationship is more complicated. The fact that the earth-current resultant varies mostly in intensity only, not in direction, makes it unlikely that the components should differ

in origin. Above all, it has been found by spherical harmonic analysis that the primary cause of the daily magnetic variations lies above the earth (e.g. in the ionosphere), and only the smaller part is due to currents within the earth. Even these currents seem to flow mostly at a considerable depth, below 200 km. From this standpoint, the earth-currents, as observed, appear as secondary effects. The induction effect
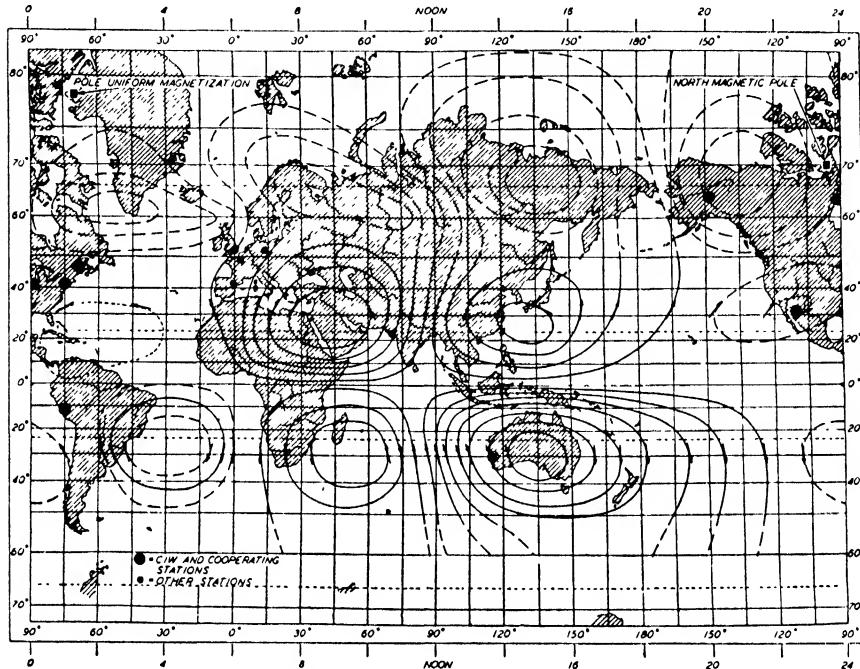
FIG. 9. The earth-current system at 6h G.M.T., derived from the observations of Gish and Rooney

of the magnetic variations must, of course, also take into account the variations of the vertical component Z (in horizontal 'loops'). In fact, on certain assumptions Chapman and Whitehead [70] computed the earth-current daily variations from the distribution of the variations of Z over the earth, and obtained good qualitative agreement between the computed and the observed daily variations of the N and E earth-currents at Ebro [71]; and the fact that the observed gradients are about five times larger than the computed gradients must not be taken too seriously, because of the high resistivity around Ebro. The fact that earth-current oscillations increase in amplitude with decreasing period is also in favour of the induction hypothesis. But a complete quantitative explanation cannot yet be given.

It has at times been suggested that earth-currents are nothing but induction effects in the wires connecting the electrodes. The wires with the earth between were considered as a current-loop in which the time-changes of the horizontal magnetic component normal to the loop
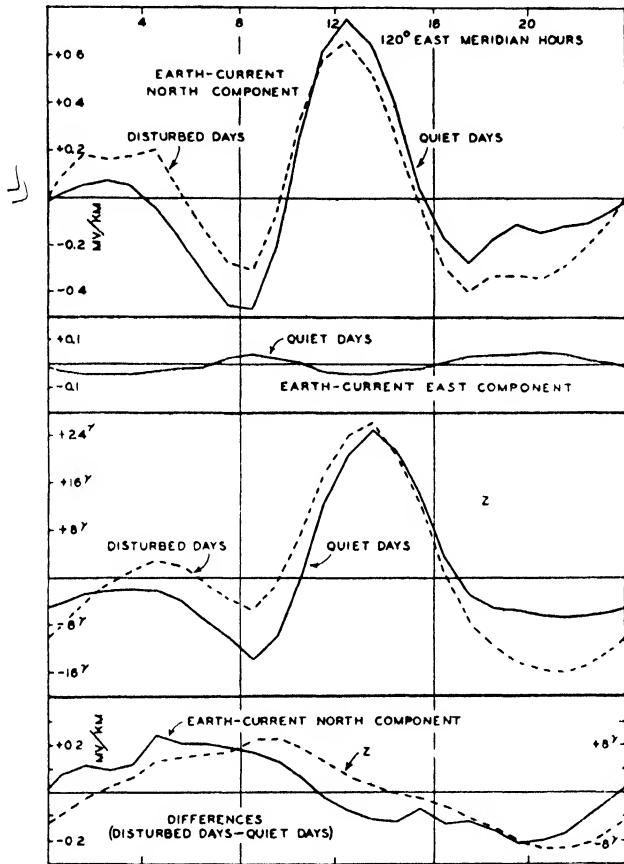
FIG. 10. Daily variation curves for the component earth-current intensities and for the magnetic vertical intensity Z, on quiet and disturbed days, at Watheroo, West Australia, 1924–8. (After Rooney)

induced a voltage, in opposite directions in the wire and in the earth. This hypothesis fails quantitatively (unless the return current in the earth is supposed to flow at considerable depths, $h = 200$ km. or more), and van Bemmelen [57] has brought another argument against it: The voltage in two parallel lines should then be twice that in a single wire, because the loop then contains two windings. But the observations on double lines give the same voltage as on single lines.
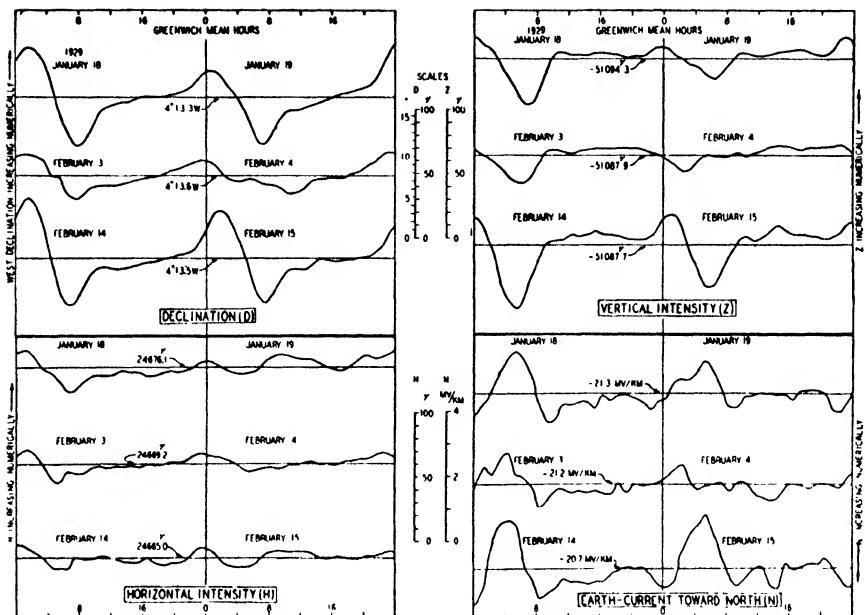
FIG. 11. The exceptionally small daily variations in geomagnetism and earth-currents, on February 3 and 4, 1929, as compared with the ordinary daily variations on two earlier and two later days, at Watheroo Observatory (Greenwich midnight is 7·7ʰ mean local time)
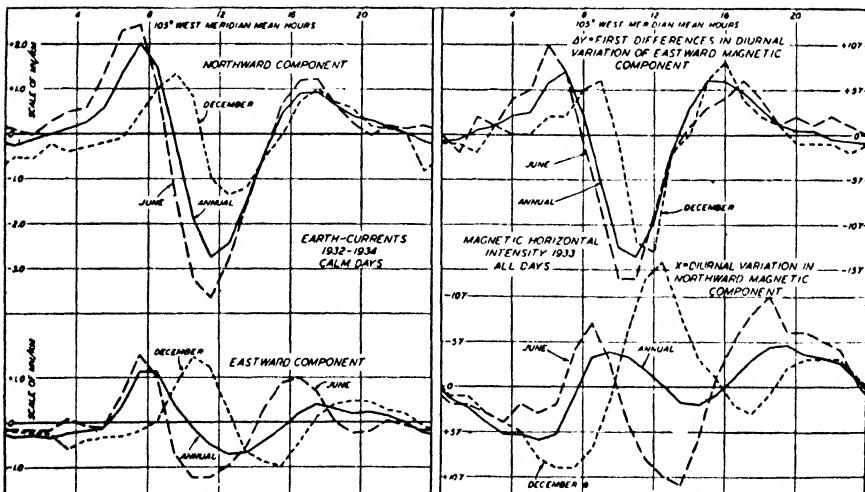
FIG. 12. Daily variation curves, showing parallel changes in the earth-current potential gradient (components northward and eastward) and the magnetic horizontal components (X, inverted curve, and changes ΔY of Y, from hour to hour), in June and December at Tucson, Arizona. (After Rooney)

13.18. Electric currents induced in moving water. Faraday, in his Bakerian Lecture of 1832, pointed out that in water, moving across the earth's magnetic field, an appreciable electromagnetic force should be induced. In his own experiments he lowered metallic electrodes from Waterloo Bridge into the tideway, but the results were inconsistent. However, electrical disturbances, which could definitely be traced to movements of masses of sea-water in the earth's magnetic
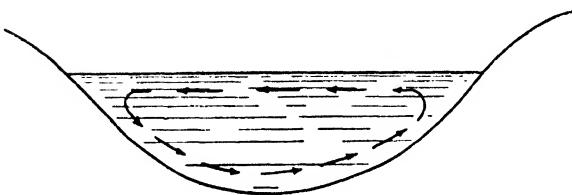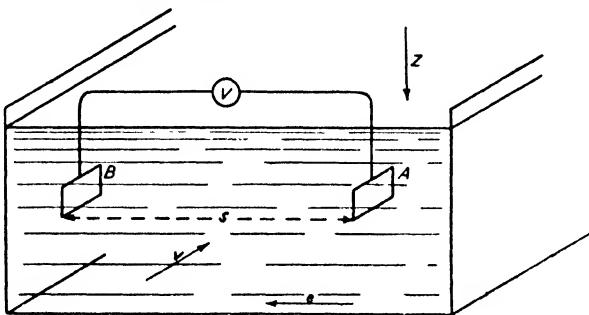
FIG. 13. Electromotive forces induced in moving water by the geomagnetic field

field, were found in sea experiments of the British Admiralty [74] in 1918, and the sea was recognized as a region of considerable electromagnetic activity.

A stream of water is, so to speak, composed of transverse filaments which in their horizontal motion cut through the lines of the earth's vertical magnetic field. The electromotive force e is induced from right to left as one faces downstream, and for a (uniform) horizontal velocity v its value in a (horizontal) filament of length s, moving across the earth's vertical component Z (Fig. 13), is given by

$$e = Zvs \text{ electromagnetic units or } Zvs \times 10^{-8} \text{ volt.} \qquad (10)$$

Only the effect of Z is considered, because the vertical electromotive components induced by X and Y were not measured. Near Dartmouth

on the English Channel $Z = 0\cdot43$ gauss; if $v = 50$ cm./sec. (about a knot), e is about 2×10^{-7} volt per cm. per knot, or 20 millivolts per km. per knot.

If the banks and bed of the channel are not conducting, no current will be produced. But if they are conducting (and that is indicated in the experiments described in § 7 (a)), a current will flow from right to left across the stream, returning through the earth. The current-density i will reach a maximum $i_{\max}$ if the resistance of the earth-return circuit is negligible compared with that of the water. In this case, taking $\rho = 21$ ohm cm. as the resistivity of sea-water,

$$i_{\max} = (e/s)/\rho = 10^{-8} \text{ amp./cm.}^2 \tag{11}$$

For actual observations electrodes must be used. Two stationary electrodes A and B may be moored on a line transverse to the water flow, at a distance s apart, and may be connected by stationary leads through a voltmeter V (Fig. 13). The electromotive force e_1 acting from A to B, recorded by V, will depend on e and on the current-density i in the water. The current i will reduce the voltage in the water to $e - i\rho s$, so that

$$e_1 = -e + i\rho s. \tag{12}$$

If $i = 0$, $e_1 = -e$; if $i = i_{\max} = e/s\rho$, $e_1 = 0$. Thus e_1 may have any value between 0 and $-e = -Zvs$, which is the maximum numerical value, corresponding to insulating banks.

If the electrodes and their connexions are allowed to *drift* with the tidal stream, an electromotive force e is induced in the connecting cable in the same manner as in the water filaments. Then the voltmeter measures

$$e_2 = e_1 + e = i\rho s = (i/i_{\max})e, \tag{13}$$

the maximum value e being obtained for a perfectly conducting earth-return circuit.

In actual experiments the velocity v diminishes towards the banks and the bottom; if the banks were non-conducting, this might lead to electrical 'convection currents' across the stream (Fig. 13, lower diagram). If the stationary electrodes are moored near the coast, currents i might be circulating which are produced by the more extensive tides in the open sea, and they might even produce effects e_1 opposite to those expected from the locally observed velocities.

The actual observations (Fig. 14) were made with silver electrodes coated electrolytically with silver chloride; they were tolerably free from polarizing effects. The distances were $s = 200$ and $2,000$ metres.
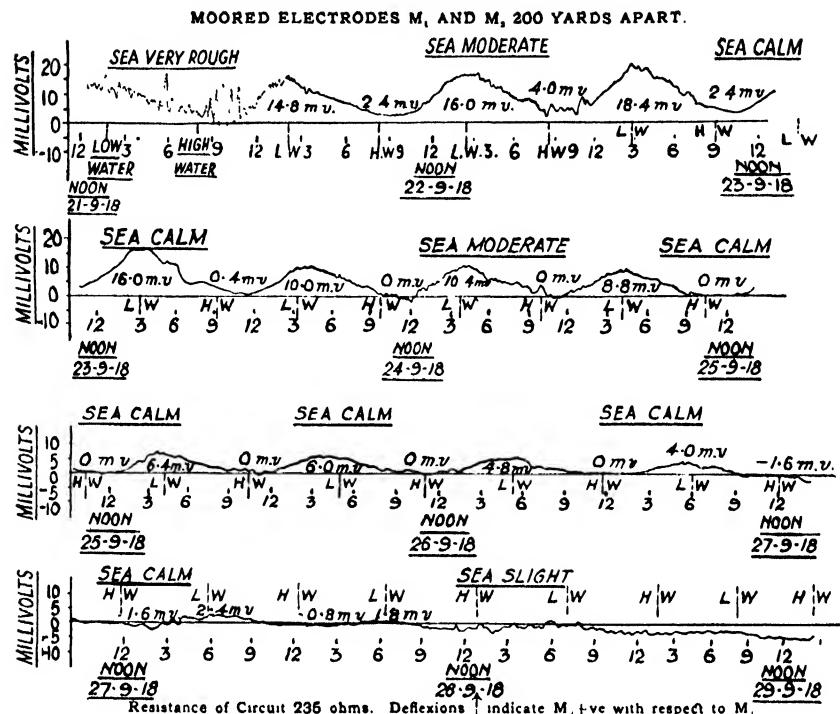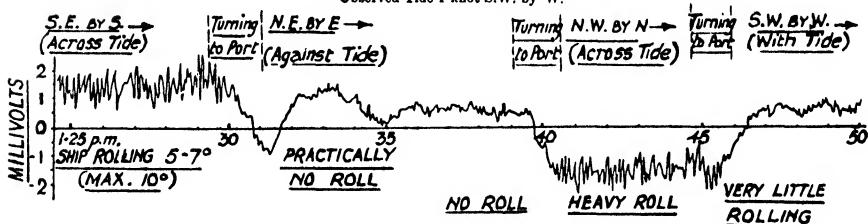
FIG. 14. Observed voltages between electrodes in sea-water.
(After Young, Gerrard, and Jevons)

The records showed curves of tidal character, the 'electromagnetic tide' diminishing markedly from spring tide (21 Sept. 1918) to neap (28 Sept.). The phase-relations suggested that in (12) the term ips plays a predominant part, i being determined much less by the local tidal flow than by the general flow in that part of the English channel. The recorded values were of the order 20 to 35 millivolts per km., corre-

sponding to tidal velocities of 1 to 1·7 knots. The 'tidal dynamo' is almost completely short-circuited by the earth, as was to be expected, that is, i is nearly i_{max}. In the absence of currents generated at larger distances, as in the Thames at London, e is therefore nearly compensated by $i\rho s$ in (12), and e_1 is small, as observed by Faraday.

Drifting electrodes, 100 metres apart, were towed in tandem by a vessel steering different courses relative to the tide. As expected, the voltage was reversed when the motion across the tide from left to right was reversed, while steering with or against the tide gave no differences.

In wave-motion caused by the wind, water filaments parallel to the wave-crests revolve in cylindrical paths passing to and fro across the magnetic lines of force. For waves propagated along the magnetic meridian the induction is associated with the total intensity of the earth's field, F; for waves propagated towards east or west, only the vertical component Z is effective. For a wave of about 10 feet height, the revolving velocity near the surface is about 150 cm./sec. or 3 knots. This induces 6 millivolts per 100 metres. Effects of this kind were observed when towing the electrodes; when the wind came at right angles to the course ('abeam') and the wave-crests were parallel to the line between the electrodes, large oscillations of the same wave-frequency, and of 1 to 1·5 millivolts amplitude, were observed; these fluctuations subsided when the waves travelled parallel to the ship, so that their crests crossed the path of the ship.

THE AURORA POLARIS

14.1. General appearance. In contrast to the magnetic phenomena, which can be perceived only with the help of instruments, the aurora is visible to the eye, and indeed often presents a striking aspect. The actual impression caused by its more magnificent displays transcends verbal descriptions and reproduction by photography or painting. But some good accounts of auroral displays will be quoted here by way of introduction.

The aurora polaris, or polar aurora, is known also as the aurora borealis, or northern lights, in the northern hemisphere, and as the aurora australis, or southern lights, in the southern hemisphere. The northern lights were known to the Greeks and Romans, and are mentioned by Aristotle, Pliny, and Seneca. We quote a passage of Seneca (*Naturales Questiones* I, 14, 15)—after Angot [1], who terms it curiously exact.

'Sometimes flames are seen in the sky, either stationary or full of movement. Several kinds are known: the abysses, when beneath a luminous crown the heavenly fire is wanting, forming as it were the circular entrance to a cavern; the tuns (*pithitae*), when a great rounded flame in the form of a barrel is seen to move from place to place, or to burn immovable; the gulfs (*chasmata*), when the heaven seems to open and to vomit flames which before were hidden in its depths. These fires present the most varied colours: some are a vivid red; others resemble a faint and dying flame; some are white; others scintillate; others finally are of an even yellow, and emit neither rays nor projections. Among these phenomena should be ranged those appearances as of the heavens on fire so often reported by historians; sometimes these fires are high enough to shine among the stars; at others, so low that they might be taken for the reflection of a distant burning homestead or city. This is what happened under Tiberius, when the cohorts hurried to the succour of the colony of Ostia, believing it to be on fire. During the greater part of the night the heaven appeared to be illuminated by a faint light resembling a thick smoke.'

In the sixteenth and seventeenth century aurorae were often confounded with comets. They were then regarded as supernatural forewarnings of horrible calamities; tails, bloody swords, armies in battle were distinguished in the sky, and caused superstitious terror in those parts of Europe where aurorae appear only occasionally.

In the northern countries aurorae were always a familiar phenomenon.

In about the year 1250 a Norwegian describes the aurora in the following remarkable passage from 'The King's Mirror' (*Konungs Skuggsjà*) [11].

'This is the nature and constitution of the northern light, that it is always the brighter, the darker the night, and it is seen only at night, never by day, and especially in profound obscurity, and rarely by moonlight. It appears like a flame of a strong fire seen from afar. Pointed shafts of unequal and very variable size dart upward into the air, so that now the one and now the other is the higher, and the light is floating like a shining blaze. So long as these rays are highest and brightest, this sparkling fire gives so much light that, out of doors, one can find one's way about and even hunt. In houses with windows it is light enough for men to see each other's faces. But this light is so variable, that it sometimes seems to grow obscure, as if a dark smoke or a thick fog is breathed on it, and soon the light seems to be stifled in this smoke and near extinct. But as soon as this fog grows less thick, the light is purged and brightens again, and sometimes it seems to shoot out big sparks, like a red-hot iron taken from the forge. As night ends and dawn approaches, the light begins to pale, and disappears when day breaks. Some people maintain that this light is a reflection of the fire which surrounds the seas of the north and of the south ; others say that it is the reflection of the sun when it is below the horizon ; for my part I think that it is produced by the ice which radiates at night the light which it has absorbed by day.'

Finally we quote the famous description by Weyprecht, who wintered with the Austrian-Hungarian Arctic Expedition in 1872–3 south of Franz-Josef's Land (78° N., 60° E.), well within the zone where aurorae appear as frequently to the north as to the south of the zenith.

'There in the south, low on the horizon, stands a faint arch of light. It looks as it were the upper limit of a dark segment of a circle ; but the stars, which shine through it in undiminished brilliancy, convince us that the darkness of the segment is a delusion produced by contrast. Gradually the arch of light grows in intensity and rises to the zenith. It is perfectly regular ; its two ends almost touch the horizon, and advance to the east and west in proportion as the arch rises. No beams are to be discovered in it, but the whole consists of an almost uniform light of a delicious tender colour. It is transparent white with a shade of light green, not unlike the pale green of a young plant which germinates in the dark. The light of the moon appears yellow contrasted with this tender colour, so pleasing to the eye and so indescribable in words, a colour which nature appears to have given only to the Polar Regions by way of compensation. The arch is broad, thrice the breadth, perhaps, of the rainbow, and its distinctly marked edges are strongly defined on the profound darkness of the Arctic heavens. The arch mounts higher and higher. An air of repose seems spread over the whole phenomenon ; only here and there a wave of light rolls slowly from one side

to the other. It begins to grow clear over the ice; some of its groups are discernible. The arch is still distant from the zenith, a second detaches itself from the dark segment, and this is gradually succeeded by others. All now rise towards the zenith; the first passes beyond it, then sinks slowly towards the northern horizon, and as it sinks loses its intensity. Arches of light are now stretched over the whole heavens; seven are apparent at the same time on the sky, though of inferior intensity. The lower they sink towards the north the paler they grow, till at last they utterly fade away. Often they all return over the zenith, and become extinct just as they came.

'It is seldom, however, that an Aurora runs a course so calm and so regular. The typical dark segment which we see mentioned in treatises on the subject in most cases does not exist. A thin bank of clouds lies on the horizon. The upper edge is illuminated; out of it is developed a band of light which expands, increases in intensity of colour, and rises to the zenith. The colour is the same as in the arch, but the intensity of the colour is stronger. The colours of the band change in a never-ceasing play, but place and form remain unaltered. The band is broad, and its intense pale green stands out with wonderful beauty on the dark background. Now the band is twisted into many convolutions, but the innermost folds are still to be seen distinctly through the others. Waves of light continually undulate rapidly through its whole extent, sometimes from right to left, sometimes from left to right. Then, again, it rolls itself up in graceful folds. It seems almost as if breezes high in the air played and sported with the broad flaming streamers, the ends of which are lost far off on the horizon. The light grows in intensity, the waves of light follow each other more rapidly, prismatic colours appear on the upper and lower edge of the band, the brilliant white of the centre is enclosed between narrow stripes of red and green. Out of the band have now grown two. The upper continually approaches the zenith, rays begin to shoot forth from it towards the point near the zenith to which the south pole of the magnetic needle, freely suspended, points. [Weyprecht calls this direction—that of $-F$—the direction of the magnetic pole—this, of course, is an uncommon usage. It is more frequently called the *magnetic zenith* direction.]

'The band has nearly reached it, and now begins a brilliant play of rays lasting for a short time, the central point of which is the magnetic pole—a sign of the intimate connexion of the whole phenomenon with the magnetic forces of the earth. Round the magnetic pole short rays flash and flare on all sides, prismatic colours are discernible on all their edges, longer and shorter rays alternate with each other, waves of light roll round it as a centre. What we see is the auroral corona, and it is almost always seen when a band passes over the magnetic pole. This peculiar phenomenon lasts but a short time. The band now lies on the northern side of the firmament, gradually it sinks, and pales as it sinks; it returns again to the south to change and play as before. So it goes on for hours, the Aurora incessantly changes place, form, and intensity. It often entirely disappears for a short time, only to

appear again suddenly, without the observers clearly perceiving how it came and where it went; simply, it is there.

'But the band is often seen in a perfectly different form. Frequently it consists of single rays, which, standing close together, point in an almost parallel direction towards the magnetic pole. These become more intensely bright with each successive wave of light; each ray appears to flash and dart continually, and their green and red edges dance up and down as the waves of light run through them. Often, again, the rays extend through the whole length of the band, and reach almost up to the magnetic pole. These are sharply marked, but lighter in colour than the band itself, and in this particular form they are at some distance from each other. Their colour is yellow, and it seems as if thousands of slender threads of gold were stretched across the firmament. A glorious veil of transparent light is spread over the starry heavens; the threads of light with which this veil is woven are distinctly marked on the dark background; its lower border is a broad intensely white band, edged with green and red, which twists and turns in constant motion. A violet-coloured auroral vapour is often seen simultaneously in different parts of the sky.

'Or, again, there has been tempestuous weather, and it is now, let us suppose, passing away. Below, on the ice, the wind has fallen; but the clouds are still driving rapidly across the sky, so that in the upper regions its force is not yet laid. Over the ice it becomes somewhat clear; behind the clouds appears an Aurora amid the darkness of the night. Stars twinkle here and there; through the opening of the clouds we see the dark firmament, and the rays of the Aurora chasing one another towards the zenith. The heavy clouds disperse, mist-like masses drive on before the wind. Fragments of the northern lights are strewn on every side; it seems as if the storm had torn the Aurora bands to tatters, and was driving them hither and thither across the sky. These threads change form and place with incredible rapidity. Here is one! lo, it is gone! Scarcely has it vanished before it appears again in another place. Through these fragments drive the waves of light: one moment they are scarcely visible, in the next they shine with intense brilliancy. But their light is no longer that glorious pale green; it is a dull yellow. It is often difficult to distinguish what is Aurora and what is vapour; the illuminated mists as they fly past are scarcely distinguishable from the auroral vapour which comes and goes on every side.

'But, again, another form. Bands of every possible form and intensity have been driving over the heavens. It is now eight o'clock at night, the hour of the greatest intensity of the northern lights. For a moment some bundles of rays are to be seen in the sky. In the south a faint, scarcely visible band lies close to the horizon. All at once it rises rapidly, and spreads east and west. The waves of light begin to dart and shoot, some rays mount towards the zenith. For a short time it remains stationary, then suddenly springs to life. The waves of light drive violently from east to west, the edges assume a deep red and green colour, and dance up and down. The rays

shoot up more rapidly, they become shorter; all rise together and approach nearer and nearer to the magnetic pole. It looks as if there were a race among the rays, and that each aspired to reach the pole first. And now the point is reached, and they shoot out on every side, to the north and the south, to the east and the west. Do the rays shoot from above downwards, or from below upwards? Who can distinguish? From the centre issues a sea of flames: is that sea red, white, or green? Who can say? It is all three colours at the same moment. The rays reach almost to the horizon: the whole sky is in flames. Nature displays before us such an exhibition of fireworks as transcends the powers of imagination to conceive. Involuntarily we listen: such a spectacle must, we think, be accompanied with sound. But unbroken stillness prevails; not the least sound strikes on the ear. Once more it becomes clear over the ice, and the whole phenomenon has disappeared with the same inconceivable rapidity with which it came, and gloomy night has again stretched her dark veil over everything. This was the Aurora of the coming storm—the Aurora in its fullest splendour. No pencil can draw it, no colours can paint it, and no words can describe it in all its magnificence.'

The aurora lasts sometimes only a few minutes, but mostly several hours or even the whole night (see § 9). Though the phenomenon may be very impressive because of its changing form, intensity, movement, and vivid colour, the absolute intensity of the light is never great. The stars always glimmer through it. The total illumination of a horizontal surface by the aurora is rarely brighter than that caused by the full moon, which may reach 0·3 lux, or 3 foot-candles. Strong aurorae give enough light to read by. The apparent surface brightness of the phenomenon, as seen in the sky, only rarely surpasses that of the moon; Störmer [16] cites an observation of a ray of that intensity. Strong aurorae could be discerned during dawn, when the sun was less than six degrees below the horizon and only a few of the brightest stars were visible [16]; but all reports of aurorae being seen in broad daylight are apparently due to confusion with cirrus-clouds.

The first successful photographs of aurorae were obtained by Brendel at Bossekop, Lapland, in 1892, with exposures of 7 or more seconds [23]. By means of cameras with cinema lenses (for instance, diameter 2·5 cm. focal distance 5 cm.) and with highly sensitive plates, strong polar lights are photographed by Störmer with exposures as short as $\frac{1}{2}$ second, and even motion pictures have been obtained by Störmer, Vegard, and others during a magnetic storm (Vegard [15]). Harang and W. Bauer [24–6], with a lens of $F:1·25$ and Agfa-Kinechrom film, obtained pictures in natural colours with 2 seconds exposure.

Aurora has no connexion with weather or atmospheric-electric phenomena such as thunderstorms; clouds interfere only in so far as they hinder direct observation. But it is often accompanied by more or less strong magnetic disturbance (see § 9). No systematic difference between the character of northern and southern lights (as seen in the earth's northern and southern hemispheres respectively) has been found.

14.2. The available observations. Reports of strong aurora are abundant, and some good collections have been published [6–12]. But the network of stations at which aurorae are systematically and continuously observed is even more defective than the network of magnetic observatories, because as yet no automatic device is available to replace actual observation. The results of observations are scattered in meteorological and magnetical publications, and there is, as yet, no one standard form of publication (§ 6). Polar expeditions generally keep auroral logs, and a continuous log for the first half of the night is published for Lerwick Observatory, Shetland Islands, in the Observatories Yearbook (Meteorological Office, London) [28]. Regular observations are also available for Sitka, Alaska [29], and in the logs of the Polar Year stations.

Most of our knowledge of the position in space of aurora is due to the Norwegians, Störmer, Vegard, Krogness and others. Their stereophotogrammetric work will be described later (§ 5).

14.3. The classification of auroral forms. Each aurora appears and changes individually, but most of them are composed of one or more principal forms; more than one such form can appear at the same time. The following classification was drawn up by a Committee of the International Geodetic and Geophysical Union, and published in an atlas [5]. The name of each form is preceded by the brief symbol to be used by auroral observers in their log books.

1. *Forms without ray-structure*

HA. Homogeneous quiet arcs. These may appear with different breadths, and when they are near the horizon the upper border is often diffuse, but the lower border is sharp; this often gives the appearance of a *dark segment* between the arc and the horizon (Plate 26 a). The lower border may be regular like a rainbow, or irregular; in the latter case it is often transformed into rays soon afterwards. Several parallel arcs may appear at the same time, extending from horizon to horizon; in double arcs, the upper one may turn round at the eastern end (in the northern hemisphere) and continue as the lower arc. Parallel arcs may

a. Single homogeneous arc (*HA*) with dark segment. Near Oslo.
Altitude 6°, towards the north

b. Homogeneous band (*HB*) in folds. Near Bossekop. Altitude 20°
towards the north

*Plates 26–30 are taken from the 48 photographs in the Photographic
Atlas of Auroral Forms*

PLATE 27

a. Diffuse luminous surface (*DS*), resembling a cloud. Bossekop.
Altitude 10°, towards the north

b. Pulsating surface (*PS*). Bossekop. Altitude 66°, towards
the west

merge together and form a large zone across the heavens. The arc may also be divided into several narrow arcs, or be split up into irregular fibres in the direction of the arc. Sometimes the arc is diffuse, the parts differing in brightness; often only parts of a complete arc may appear, extending from the horizon to a certain height, or isolated parts extending neither to the east nor to the west horizon.

HB. Homogeneous bands. These are less regularly shaped, and more rapidly moving, than the arcs. The lower border is often sharp, but irregular in contour. Semi-circular or ellipsoidal shapes occur, and appear more luminous where they are seen tangentially; the band may also show one or more folds (Plate 26 *b*). The breadth may vary from a very narrow band to a band so large that it resembles a hanging curtain; these broad bands turn very often into bands with ray-structure.

PA. Pulsating arcs. Whole arcs or also parts of an arc may flash up and disappear rhythmically with a period of several seconds; this form often stands quite isolated in the sky.

DS. Diffuse luminous surfaces. These consist of a diffuse veil or glow over great parts of the sky, with indistinct boundaries, or residual luminosities which resemble clouds (Plate 27 *a*); they often appear after intense displays of rays and curtains. Sometimes large areas of the sky may be coloured by a diffuse violet and red light.

PS. Pulsating surfaces. These are diffuse patches of light which appear and disappear at the same place and in the same irregular shape (Plate 27 *b*) with a period of several seconds. Near the *magnetic zenith* (that is, the point or direction in the sky along, or opposite to, the direction of the magnetic field vector $\pm F$) such patches may be sharper.

These forms often appear in connexion with flaming aurora.

G. A *feeble glow*, near the horizon, resembling dawn, of a white or reddish colour. It may often be the upper part of an arc whose lower border is underneath the horizon.

2. *Forms with ray-structure*

These forms are made up of short or long rays, which may be arranged in different ways. All auroral rays seem to converge towards the magnetic zenith, as if they were forced to take the direction of the earth's lines of magnetic force, parallel to the axis of a freely-poised magnetic needle (a dip-needle); along the northern auroral zone the magnetic zenith is situated somewhat to the south of the true vertical zenith.

RA. *Arcs with ray-structure.* These may develop from homogeneous arcs, which may become sharp and luminous along the lower border, and then very rapidly change into an arc of rays (Plate 28 *a*). (See also Fig. 2 *a*, and Plate 31 *a*, § 5.)

RB. *Bands with ray-structure.* These resemble the homogeneous bands, but are constituted of a series of rays, either close together or more scattered (Plate 28 *b*). Often several parallel bands appear. A band near the magnetic zenith may have the form of a corona (Plate 29 *a*).

D. *Draperies.* These consist of one or more bands with very long rays, which appear like a curtain or drapery. The lower border is often the most luminous. Near the magnetic zenith the perspective may give to the curtain a fanlike form (Plate 29 *b*). (See also Fig. 2 *b* and Plate 31 *b*, § 5.)

R. *Rays* may be narrow or broad, short or long; they may appear singly, or in great bundles or masses, resembling curtains (Plate 30 *a*).

C. *Corona.* Rays near the magnetic zenith converge to this point and may form a corona (Plate 30 *b*), complete or incomplete. Bands or draperies near the magnetic zenith may also form a corona.

3. *Flaming Aurora*

F. A characteristic, rapidly-moving form. It consists of strong waves of light which rapidly move upwards one after another in the direction of the magnetic zenith. The waves may have the form of detached arcs which move upwards normally to the direction of the arcs, or they may be compared with invisible waves, which in their passage illuminate broad rays and patches, which appear and disappear rhythmically when the waves pass them.

This form often follows strong displays of rays and curtains, and frequently precedes the formation of a corona.

The colour of aurorae is mostly greenish yellow; sometimes it is bluish-white, like the light of an electric arc-lamp, or red, and occasionally dark red; also reddish-white and violet-grey colours have been observed. According to Vegard, draperies are usually greenish-yellow, and veined with reddish-blue streamers. The lower borders of draperies are sometimes dark red. Rays are observed in all the usual auroral colours. Störmer [16] has observed rays which, without altering either in form or position, changed their colour successively from bluish-green to reddish-violet along their whole extent.

PLATE 28

a. Arc with ray-structure (*RA*). Bossekop. Altitude 17°, towards the north

b. Band with ray-structure (*RB*). Bossekop. Altitude 14° towards the west

PLATE 29

a. Band with ray-structure (*RB*) near the magnetic zenith, forming a corona. Oslo

b. Drapery (*D*) near the magnetic zenith. Oslo

Vegard [30] has described cases in which different forms were gradually developed and transformed into one another. Also, after a a few minutes of comparatively fixed position, the aurora may within a few seconds change its form and also appear in quite different areas of the sky.

The homogeneous arc is the most frequent form, and flaming aurora is the most impressive form. Towards the end of one strong aurora Störmer [16] saw repeatedly flames shooting up to the magnetic zenith in large and diffuse rays, renewed at every instant for at least 15 minutes. He had the distinct impression of electrical discharges directed outwards into space.

14.4. The height of the aurora—early observations. Observation from a single station does not enable the height of an aurora to be determined, though it may suggest that the aurora is above the clouds which may partly veil it. But from values of the azimuth and angular height of identical points of the aurora, as seen from two stations at a sufficient distance apart, the absolute height and position in space can be calculated by trigonometrical formulae, much in the same way as the height of a mountain is measured in geodetic surveys. The accuracy depends on the parallax, that is, the angle between the lines of sight from the two stations; at the ends of a 'base-line' of 20 km. length, an object at 600 km. distance has a parallax of 2°, or four times the diameter of the moon.

This method seems first to have been applied by Mairan [6] in 1726. Contrary to his expectation he always found considerable heights, more than 100 leagues. Similar observations by others also gave great altitudes. But in many cases the method failed completely; for instance, an aurora seen in the north was measured at one station as being at a lower apparent height than at a more southerly station [1]. These failures in the practical application of a simple theory are due to various reasons. One main reason is that many aurorae are more or less diffuse, often without distinct marks which could be clearly identified by two observers at a distance from one another; even the borders of luminous areas, as seen from two standpoints, may not correspond to the same points in space, if the aurora has a certain depth and is not confined to a thin film. Furthermore, the condition that the observations of mobile forms must be simultaneous within a few seconds presented a difficulty which was only solved after telephonic communication became possible.

A quiet arc near the zenith gives the best opportunity for measurements. Cavendish [31] used observations of an arc taken from two

stations on February 23, 1784. At Cambridge the position of the arc in the sky had been described by the stars lying along it; from these indications the arc was found to follow a great circle which passed only 1° north of the zenith, and the western end had the direction 18° south of west. If a line is drawn through Cambridge parallel to the arc, the other station (Kimbolton), where the arc had been found, by the use of a quadrant, to pass 11° to the south of the zenith, is situated 23·8 km. to the north of it. Hence (cf. Fig. 1) the height, h, of the arc is easily found from the equation

$$h \tan 1° + h \tan 11° = 23\cdot8 \text{ km.,}$$

which gives $h = 112$ km. This value is uncertain in so far as the observations at Cambridge can also be interpreted to give $h = 84$ km.; but both limits agree quite well with modern measurements.

Among the other early measurements, we may mention that of the chemist Dalton [6 a] (163 km.), and those of Bravais at Bossekop, Lapland, who found 50 to 100 km. Floegel [32] analysed the observations of the great aurora of October 25, 1870; he found for the lower ends of the rays heights of 100 to 300 km., and for their summits often more than 500 km., and sometimes even 750 km. These observations retain their value, because they were taken during one of the strongest displays in history (see § 9).
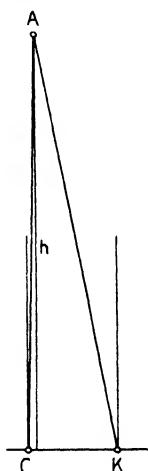
FIG. 1. The calculation of the height of an auroral arc by Henry Cavendish, from observations at Cambridge (C) and Kimbolton (K)

14.5. Photographic measurements. Our present knowledge about the position of the aurora in space is almost entirely due to the photographic measurements by the Norwegians, Störmer, Vegard, and their collaborators. The aurora was photographed simultaneously from two or more stations; these were situated at mutual distances from 15 to about 400 km., and were connected by telephone, so that the observers could direct their cameras towards the same region of the sky and expose simultaneously. The pictures of the stars on the plate fix the astronomical orientation; the shift of the aurora relative to the background of the stars appearing on the different photographs is used to calculate the position of the aurora. This trigonometrical calculation is abbreviated by the use of graphical and instrumental devices. Suitable star-charts for use in such work have been published in the supplement to the *Atlas of Auroral Forms* [5].

PLATE 30

a. Mass of rays (*R*). Oslo. Altitude 34°, towards the north-east

b. Corona (*C*) at the magnetic zenith. Oslo

PLATE 31
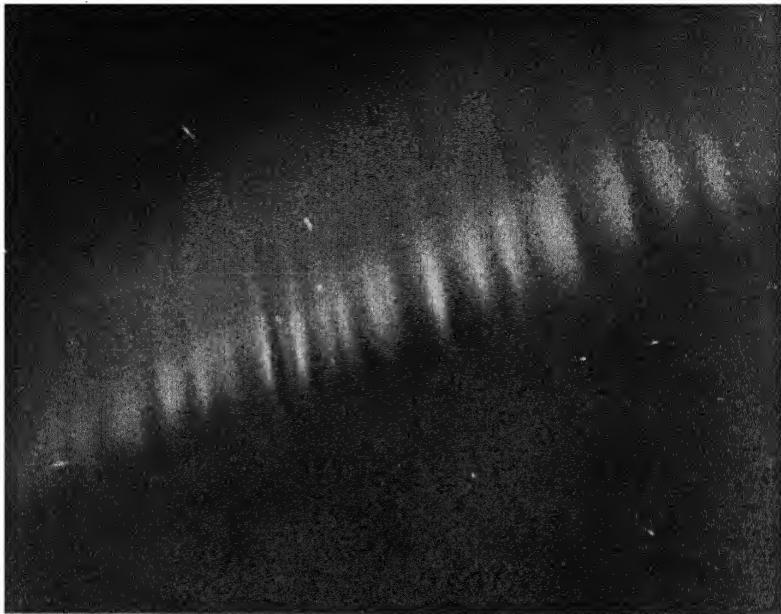
a. An auroral arc of 1936 March 24 photographed at 22^h $24\frac{1}{2}^m$ G.M.T. (Störmer)

b. An auroral drapery of 1936 January 24 photographed at 22^h 29^m 29^s G.M.T.
(Störmer)

Long series of measurements have been taken at two places in northern Norway, by Störmer [14] in 1910, at Bossekop (70° N., 23° E.; angular distance 23·5° from magnetic axis), and near by, at Haldde Observatory, by Vegard and Krogness in 1912 and 1914 [15]. Since 1911 Störmer has observed in southern Norway, near Oslo (59·9° N., 10·7° E.; angular distance from magnetic axis 30·0°), in conjunction with a network of auroral stations which he has organized and directs. He has published his results up to 1922 [16] and also accounts of many selected later observations [34–7]. Recently more photographic measurements of auroral heights have been made in Alaska by Fuller [19], at the British Polar Year station Fort Rae by Stagg [G 94], and at other stations.

The number of satisfactory pairs of plates taken in Norway was already, in 1922, well above 1,000, and the number of points measured is a multiple of this; but it must be remembered that many plates were taken for the same aurora; for Oslo, 1911–22, all the material used referred to twenty-two nights. As to the accuracy, Störmer states that, with three base-lines as long as 70 km., the heights computed for each pair of stations differed by about 2 per cent.; this corresponds to an error in the parallax of about 1°. The reduction of the plates is laborious in spite of ingenious graphical methods developed in Norway; other methods have been proposed, but not yet applied [35 a].

Figs. 2 a, b show the geographical positions, in plan, of the two aurorae illustrated in Plate 31 a, b, which occurred on (a) 1936 March 24, (b) 1936 January 24. The former is an auroral arc with ray structure (RA, in the nomenclature of § 2), and the latter is an auroral curtain or drapery (D, § 2). Plate 31 b is one of 260 photographs taken on the night of 1936 January 24/25, at Störmer's auroral stations; these 260 photographs included 90 exposures made at only one station, 60 sets from two stations simultaneously, 14 sets from three, and 2 sets from four stations simultaneously. Plate 31 b was taken at Lillehammer, whose geographical situation is shown on Fig. 2 b (Li); simultaneous photographs were taken at Kongsberg (K, on Fig. 2 b), Oslo, and Oscarsberg. By means of the photographs at the first two stations (Li and K) the heights and geographical positions of several points on the auroral drapery, identifiable on both photographs, were found: these are shown on Fig. 2 b, and lie on the two lines marked 82; they refer to the border of the aurora which appears lowermost on Plate 31 b, and to the lower border of the first fold of the drapery, on the left; the latter border is shown on Fig. 2 b by the line of points 82 (1, 2, 3, a, 4);
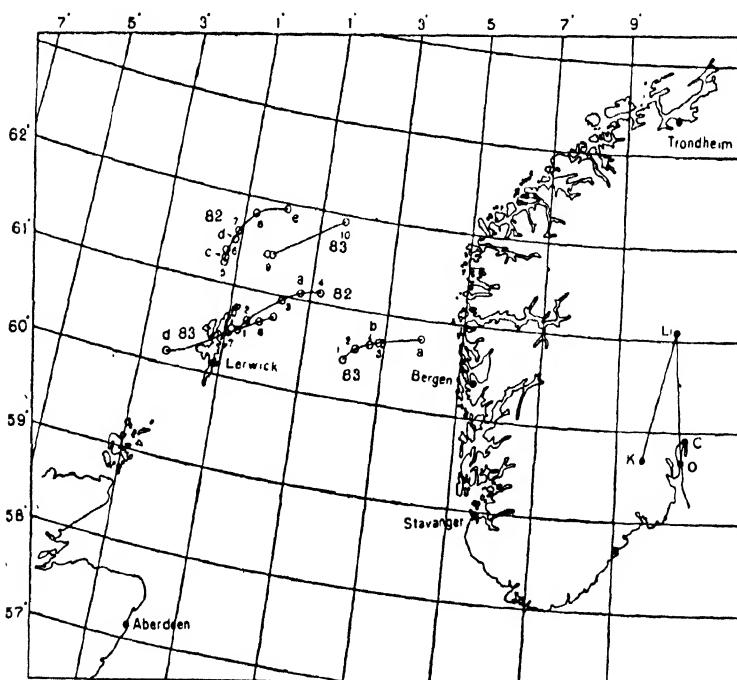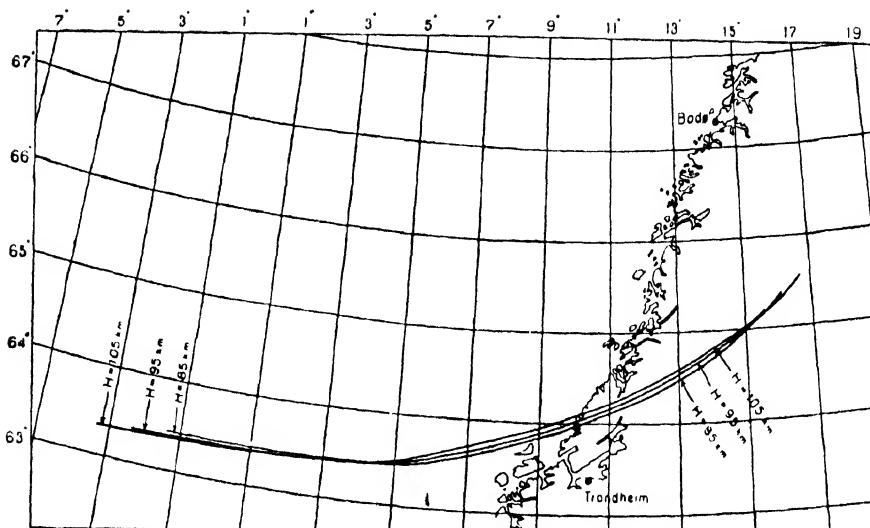
FIG. 2 *a* and *b*. Geographical positions of the auroral arc of 1936 March 24 (Fig. 2 *a*), and of the auroral curtain of 1936 January 24 (Fig. 2 *b*). (After Störmer)

the former is the other line 82 on Fig. 2 b. The drapery thus lay over the sea to the north of the Shetland Isles. Nineteen seconds later further photographs of this drapery were tàken, and though its general form was not much changed, the geographical positions of its lower borders were found to have shifted quite considerably, to the lines marked 83 on Fig. 2 b. The lower borders seem also to have risen during the 19 seconds, their heights in Plate 31 b ranging from 92 to 104 km., whereas 19 seconds later they ranged from 102 to 126 km.

Plate 31 a was taken at Trondheim, and as it was cloudy, at the time of exposure, at Störmer's other auroral stations, no simultaneous photograph could be taken elsewhere. In such cases it is often possible to determine the geographical position of an auroral arc from a photo-graph at only one (suitably placed) station, by assuming a reasonable value (guided by knowledge of other cases where real height-measure-ments have been possible) for the height H of the lower border. In the present instance the position was calculated on three assumptions as to H, namely, $H = 85$ km., $H = 95$ km., $H = 105$ km. As Fig. 2 a shows, all these lead to substantially the same estimate of the geo-graphical position; the arc lies partly over northern Norway, and extends out to sea.

Fig. 3 shows the distribution of the heights of the lower limits of various auroral forms; the curves are smoothed because of the un-certainty in the measurements. The main feature is the existence of a height-interval of maximum frequency for the lower border, situated between 106 to 107 km. in northern and 103 to 104 km. in southern Norway. Though the height of maximum frequency thus appears to lie lower in the south, the individual heights range rather widely, and it is not certain that there is a true geographical difference between the mean heights at the two places; moreover, the observations in northern Norway were taken during a sunspot-minimum, whereas the measure-ments at Oslo include some for the greatest aurorae ever seen. It is doubtful whether the undulations in the curve of frequencies are physically significant, especially the secondary maxima and the apparent split in the main maximum.

The averages of the heights for the lower limit of the various forms, taken separately, differ only by a few kilometres. Störmer considers it certain that draperies sometimes come down as low as 80 or 84 km. In February 1929 six photographs of intense short draperies were obtained with a base-length of 66 km., giving parallaxes of about 10′; the height of the lower border was 82 km. [35]. Harang and Bauer

found a height of 65 km., in March 1932, for the red lower border of a quiet arc [25]. The lower limit of draperies and arcs is generally not above 150 km., though lower borders up to 200 km. have been observed
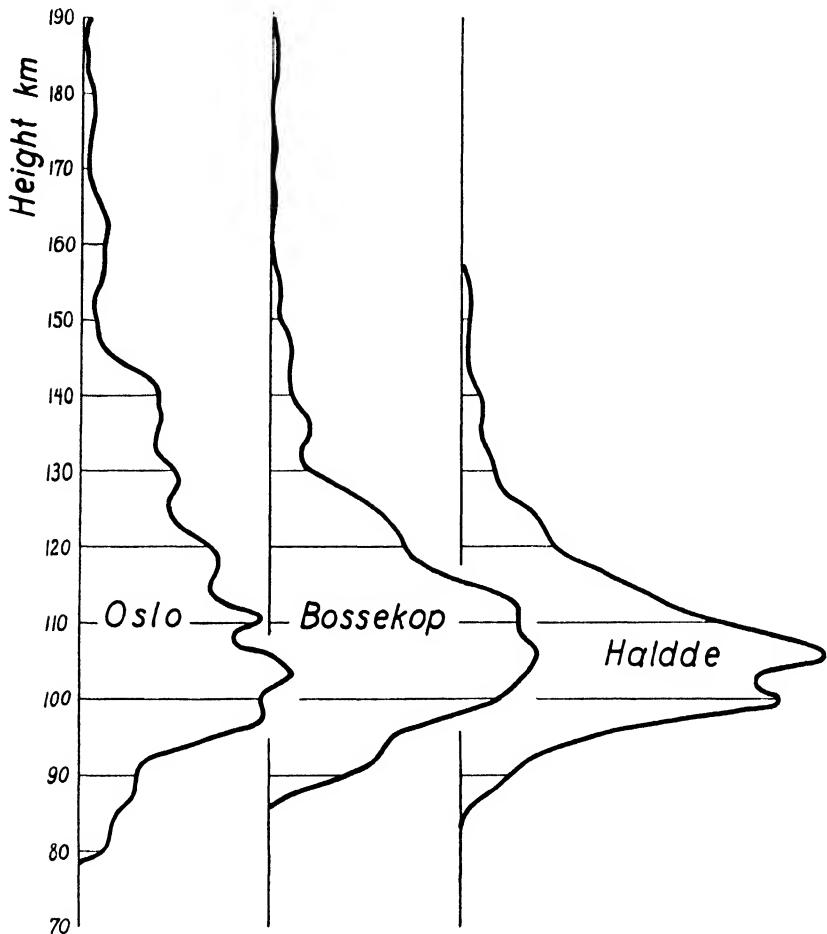
FIG 3. Diagrams showing the relative frequency of occurrence of the lower limits of aurorae at different heights, at three Norwegian stations. (After Störmer)

in exceptional cases. The vertical extent of the most luminous part, near the lower border, is generally 10 to 30 km.

The upper limits, however, vary much more. Homogeneous arcs do not often reach heights above about 150 km.; but rays extend to extreme heights, especially during intense displays. At Oslo, Störmer found more than 60 rays higher than 400 km. On March 22 and 23, 1920, rays were observed which reached up to 800 km.; even their bases were very

high, 400 km. in one case. On September 8, 1926, an auroral curtain of violet-grey colour, situated over the Shetland Islands, was observed from Oslo; its height was 300–500 km.; later it was transformed into a diffuse mass which reached up to 1,000 km. height.
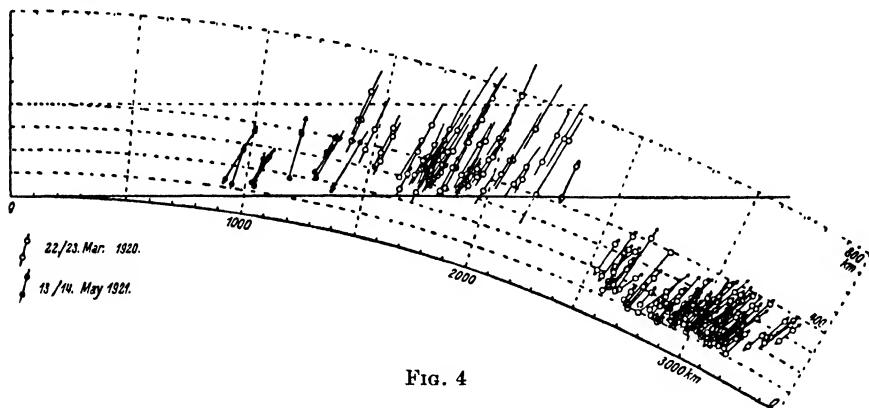
FIG. 4
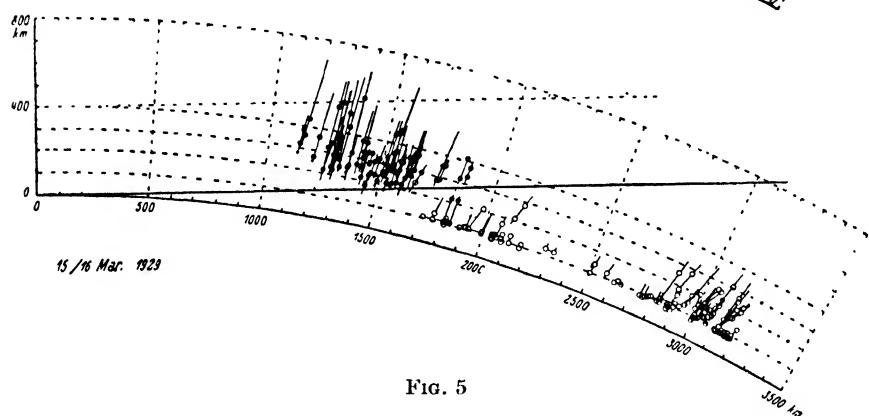
FIG. 5

FIGS. 4 and 5. Diagrams illustrating the heights, and the positions relative to the earth's shadow-boundary, of auroral rays observed by Störmer: showing the exceptional heights attained by some sunlit aurorae

Störmer [34] discovered that most rays of extreme height (above 400 km.) are situated in the sunlit part of the atmosphere (Figs. 4, 5). These rays occur over parts of the earth where the sun is as low as 20° below the horizon, and the atmosphere is in the earth's shadow up to 400 km. height; farther away from the line of sunset only aurorae of ordinary heights were seen. In some cases rays were measured which consisted of two luminous parts, one situated in the sunlight and another, along the continuation of the same line, in darkness; for instance, parts of a ray were visible above 300 km. and below 220 km.,

while the boundary between the sunlit and the dark atmosphere lay at about 275 km. (Fig. 5). The upper parts generally fade away towards great heights. The photographic picture of an aurora may differ greatly from the visual picture; Störmer has photographed ultra-violet rays which could not be detected by the eye, and infra-red photographs of the auroral spectrum have been obtained at Tromsö (in 1932) [24].

The width of auroral rays is generally several kilometres; but some rays of less than 1 km. diameter have been photographed ([16], p. 37) which, as seen from about 300 km. distance, appeared narrower than one-third of the moon's apparent diameter.

In Fig. 6 the horizontal projections of homogeneous arcs as photographed from Oslo and from Bossekop are plotted and compared with the (broken) lines of equal distance θ from the pole B of the homogeneous magnetization (here assumed in 78·5° N., 68·6° W.). The arcs nearly coincide with these circles of magnetic latitude; however, there is a small systematic deviation (10° at Bossekop according to Vegard) so that the western ends of the arcs are nearer to the pole than the eastern ends. But this deviation is not big enough to make the arcs perpendicular to the direction of declination, for the line of zero declination crosses Norway, where the arcs are never true east–west. Vegard [30] found, from a discussion of the arcs observed at many stations, that they lie more closely along the perpendicular to the direction towards the magnetic axis than along the perpendicular to the magnetic declination-needle. This was confirmed by Sverdrup from observations north of Siberia (see p. 466).

It has already been mentioned that the rays are parallel to the direction of the magnetic field-vector, as indicated by a dip-needle. On the average, the radiation-point of the rays, as observed during coronae in Oslo, is only one degree below this magnetic zenith, which over the northern hemisphere is of course situated in the southern part of the sky. Reimann and Galle [38] tried to use this fact in order to find the height of the aurora from observations at a single station P: an aurora seen in the southern sky is situated in a region where the inclination is less than at the observing station; thus if the radiation-point measured at P gives the inclination at the place Q where the aurora is actually situated, the horizontal distance of Q can be inferred, and the angular altitude of the aurora as observed at P then gives its height. But this method is rough, because the differences in inclination are small and also because the magnetic field is disturbed during aurora. An attempt [16] to coordinate the observed positions of the radiation points with
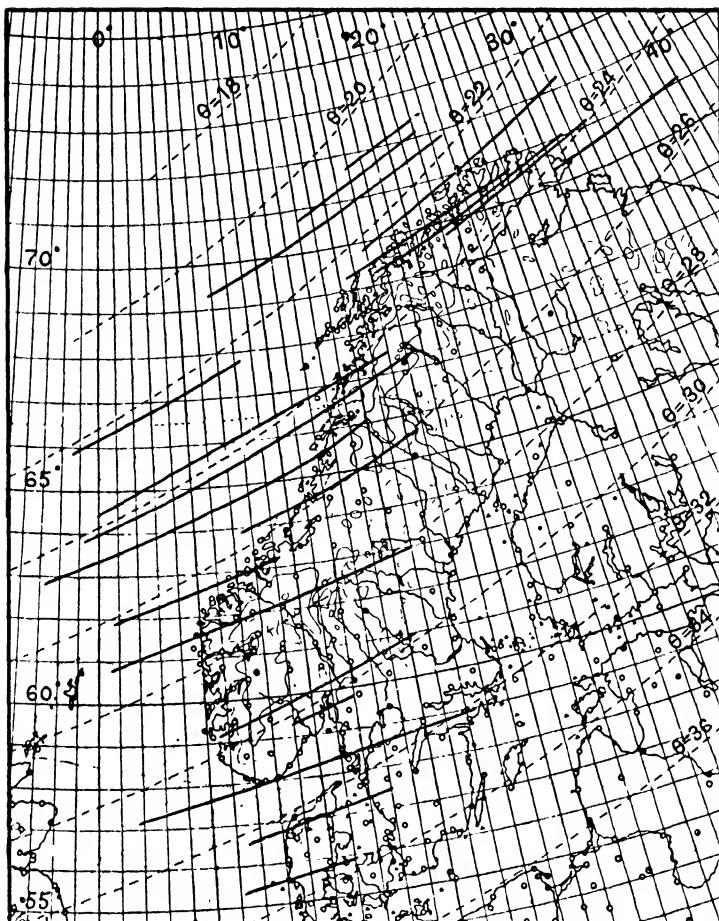
FIG. 6. Showing the geographical position and direction (in plan) of a number of homogeneous auroral arcs observed by Störmer. Their direction agrees approximately with that of the circles (broken lines) of geomagnetic latitude (θ = angular distance from the north pole of the geomagnetic axis)

simultaneous magnetic records did not lead to satisfactory results (see also § 8).

14.6. Programmes of auroral observation. The most complete and reliable results are given by simultaneous photography from two stations at a distance of more than 25 km. Various alternative programmes [5] have been planned for different circumstances and different amounts of available assistance, especially for the recent International Polar Year; for instance, if no photographic apparatus is available,

3595.30 H h

sketches of the aurora may be drawn on simple star-maps; even notes about the situation of the aurora relative to the stars may be useful.

Besides these observations, which are intended to give information about the situation of single phenomena in the atmosphere, other, more simple observations may be taken at fixed Greenwich hours, in order to obtain a statistical survey of the frequency of aurorae and of their different forms as functions of the time and of geographical position (La Cour [43 a]); during the Polar Year simple spectroscopes for detecting aurora by means of the green line were constructed and distributed [76].

Sverdrup [18] determined the position of homogeneous arcs by photographs from a single station, on the *Maud* Expedition, north of Siberia; the calculation was possible if he assumed for the lower border the same value as found in northern Norway, namely, about 110 km. In this way it was shown that the directions of the arcs coincided rather well with the circles round the magnetic axis, whereas rough eye-observations had led to an azimuth differing by 20°, and to the erroneous assumption that the arcs were perpendicular to the local direction of declination.

14.7. Low aurora and auroral sounds. In spite of the fact that the photographic observations in Norway always give heights above 80 km., many observers report that they have seen aurorae below the clouds, or near the ground, or between the observer and a mountain. The following excerpts [40] from journals of members of Greely's expedition to Lady Franklin Bay (1882–3) are characteristic:

November 17, 1882. 'The aurora of this morning was very low, and we are, I think, the only party that ever could say that we were in the midst of electric light. In fact its alarming close proximity scared one of our members considerably. . . . No noise of any kind was audible.'

'Israel thinks that at times the aurora could not have been more than 100 feet from the earth. . . . The curtain at one time appeared so near above their heads that Gardiner and Israel speak of having unconsciously dodged to avoid it.'

'One person was so much affected by the display at its grandest moments that he lowered his head and put his hands up as though to ward off a blow.'

Though most observers are impressed by the absolute silence even of strong displays, others report 'swishing' or 'rustling' sounds and even odours. But Simpson [41] concluded that the impressions of low aurora were illusions, caused by fog or mist illuminated from above. The direct relations between cirrus clouds and aurorae, which even Angot [1, p. 108 f.] did not doubt, now appear dubious, because the modern

height-measurements have shown a difference of more than 60 km. between the altitudes of clouds and aurora.

Sverdrup ascribes the swishing sound to the friction of ice crystals in the observer's freezing breath. There is, however, new evidence of sounds connected with strong aurora, heard by trained observers [37].

Chapman [39], in a review of a collection of observations of low and audible aurora, does not yet discard their possibility entirely, though no physical explanation can be given at present, and, especially, no connexion between aurora and atmospheric potential gradient or other electric phenomena near the ground has hitherto been found. Some observations are perhaps explained by a casual coincidence of aurorae with Saint Elmo's Fire [1, p. 116 f.]; Lemström [42] produced it artificially on mountains from grounded wires with points.

Apparent relations between aurorae and the topography of the land have been reported, but must be regarded with reserve. Mawson's observations in the Antarctic that a curtain, if partly hidden by a mountain, seemed to follow its contours, may have been a distortion caused by the higher density and refractive power of the cold air near the ground, just as is observed when the sun is near the horizon.

14.8. Geographical distribution. In non-polar latitudes aurorae generally appear on the poleward side of the zenith; stations situated nearer to the pole observe them brighter and nearer to their zenith. From the geometrical conditions as shown in Figs. 4, 5, it follows that the lower border of a homogeneous arc, at 100 km. altitude, can be seen from as far away as 1,000 km., and high rays can be seen from a distance of 2,000 km. and more. This is mentioned because summaries of the geographical distribution of aurorae always assign an aurora to all the stations whence it has been seen, perhaps near the horizon, and not only to the place where it was overhead.

From his catalogue of aurorae [8] Fritz [2] computed, for a number of stations, the average relative frequency of nights with aurora, expressing it as M nights per year. Since the stations did not all observe over the same periods, a reduction to a common epoch had to be applied in order to eliminate the year-to-year variations in the frequency of aurorae. This was done in a way which is usual in meteorology; the series of all observations in Central Europe, between 46° and 55° latitude, from 1700 to 1872, was chosen as a basis; 4,834 aurorae were seen in these 173 years, that is, an average of 28 aurorae per year. It was assumed that the number of aurorae seen at a single station in a certain interval is proportional to the number seen in Central Europe.

For instance in Christiania (Oslo), observations were taken from 1837 to 1870, and 965 aurorae were seen. In the same interval, 1,149 aurora were seen in Central Europe; therefore the average relative frequency, M, in Christiania is found from the proportion $965/1,149 = M/28$, which gives $M = 24$. It is clear that this method can give only rough results; the main difficulty is the great variety in the attention of the observers, since their interest, generally stimulated by a great display, often subsided afterwards.

Fritz drew a chart (Fig. 7) showing lines of equal auroral frequency, M; he called these lines *isochasms*. Since then (1881) no attempt has been made to revise this chart by means of the many observations subsequently collected; there is much need of such a revision, but Fritz's chart is an admirable piece of work which has retained its value for nearly sixty years.

Fig. 7 shows that the lines, beginning with $M = 0.1$ (one aurora in about ten years), do not follow the circles of latitude, but are more or less concentric about a point which may be regarded as identical with the pole of homogeneous magnetization at 78·5° N., 69° W. Several reports of observers in the north polar zone convinced Fritz that the frequency of aurorae does not increase towards a single polar maximum, but there is an oval or circular *zone of maximum frequency*. The average radius of this zone is 23°. Close inside this line he drew another line of neutral direction of visibility (dotted in Fig. 7), connecting the points at which aurorae appear equally often in the northern and the southern sky.

From observations on the drift of the *Maud* north of Siberia, 1918–25, Sverdrup [17] concluded that in longitude 160° east of Greenwich the difference between the frequencies of occurrence in the northern and in the southern segments of the sky disappears between latitudes 77° and 78°; he identified this line of neutral direction of visibility with the zone of maximum frequency. This location is about 23° distant from the magnetic axis. Near the maximum zone he found that mobile auroral forms predominate, but that farther to the south quiet forms are more frequent; in 70·7° N., 162·4° E., only about 7° south of the zone of maximum frequency, aurorae were most often seen as a low arc in the northern sky.

The sharp decrease of auroral frequency outside the zone of maximum frequency is borne out by many observations. For instance, of 107 strong aurorae seen at Trondheim during the years 1891–1906, only 28 were seen simultaneously at Oslo, and 13 in the Netherlands; the angular distances of these three places from the magnetic axis of the earth are 30·6°, 33·7°, and 36°, respectively (Röstad [43]).

Observations of the southern lights are scarce because of the small
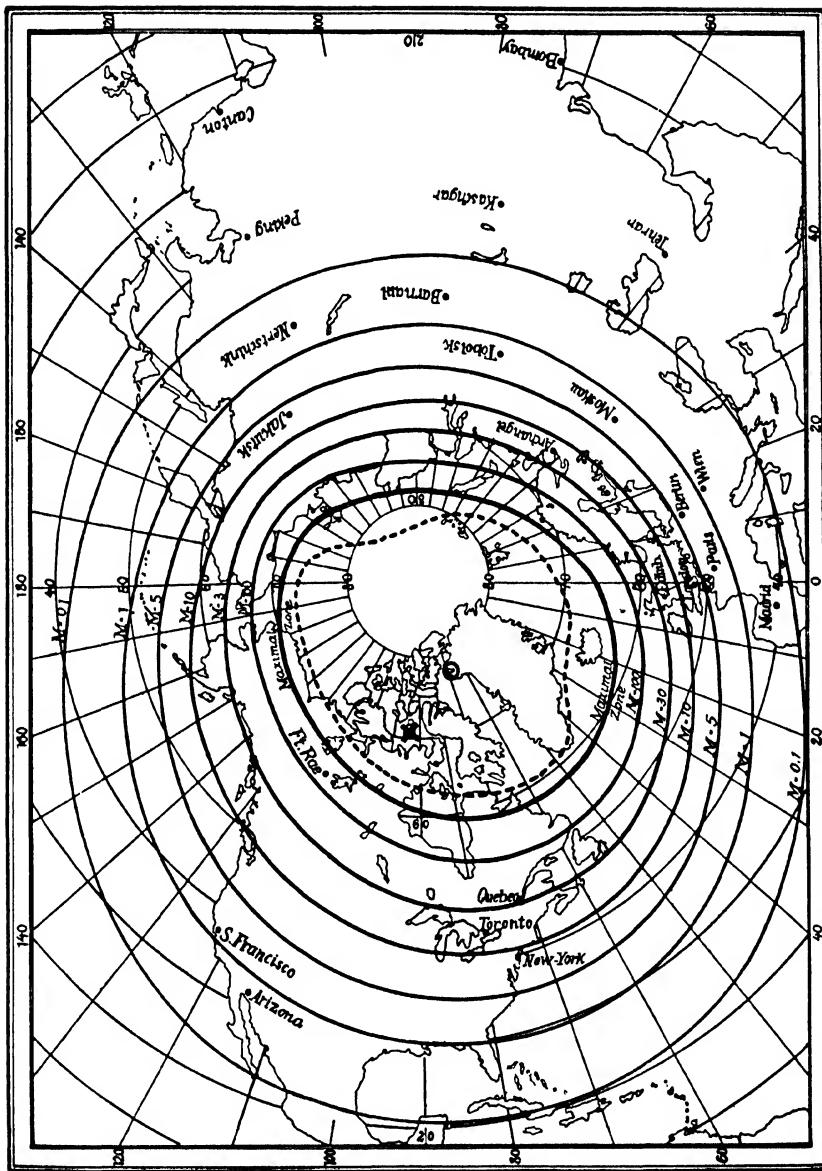
FIG. 7. The distribution of isochasms, or lines of equal auroral frequency, in the northern hemisphere, according to Fritz

number of antarctic land stations. Boller [9] compiled a catalogue of 791 days of southern lights, and drew two very conjectural zones of maximum auroral frequency in the Antarctic. But one feature of his diagram is quite analogous to Fritz's chart, namely, that the frequency of aurorae in Australia and over the Indian Ocean (at longitudes around
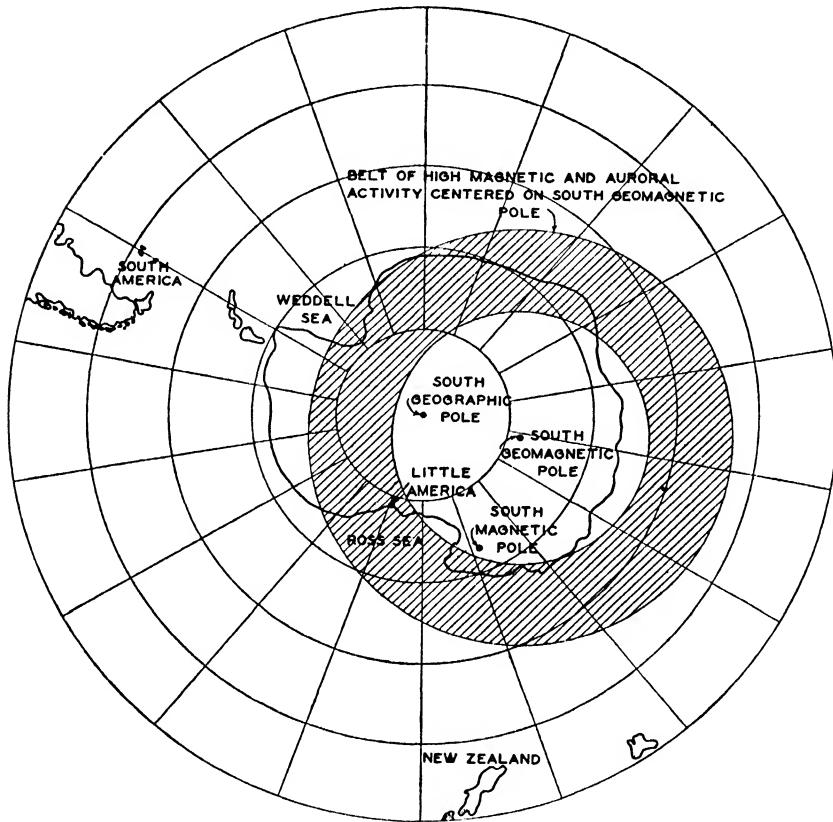
FIG. 8. The southern auroral zone, according to Davies

110° E.) is much higher than in the same latitudes of the South American quadrant (at longitudes 70° W.). This is due to the situation of the southern pole of the earth's magnetic axis, in 78·5° S., 111° E.

It will be seen later (§ 9) that the zone of maximum auroral frequency is more or less a statistical conception, since the observed geographical distribution of the single displays depends very much on their intensity; the greater the aurora and the accompanying magnetic disturbances, the more it extends from the pole.

Fig. 8 gives the southern auroral zone, after Davies [45 a].

14.9. Periods, and connexion with magnetic disturbances.
Close connexions between aurorae and geomagnetism have already been
indicated, concerning the directions of arcs and rays as well as their
average geographical distribution. Much earlier, in 1716, Halley found
that aurora is accompanied by magnetic disturbances. The correspond-
ing abnormal earth-currents were noticed later, after the construction
of telegraphic lines; in fact these phenomena are so closely connected
that Störmer has been informed by the Norwegian telegraphic offices,
when they observed disturbances due to such currents, that it seemed
promising to make ready for auroral observations in the evening.

At Lerwick Observatory (Shetland Islands), since 1924, the sky is
inspected regularly during the eight months September–April, at
intervals of 15 to 30 minutes from 19^h to 23^h G.M.T. A summary
of the five observation-periods 1924–9 has been given by Lee [28]:
the meteorological conditions, and the absence of strong moonlight,
were favourable for the observation of aurorae on 30 per cent. of
the evenings; some form of aurora was actually seen to occur on 59 per
cent. of these favourable evenings, and forms with ray-structure on
18 per cent.

By analogy with the daily magnetic character-figures, derived from
the magnetic records at Lerwick, Lee derived auroral character-figures
for each evening, defined as follows:

0 = no aurora observed, though at least three hours were favourable
for observation;

1 = auroral forms occurred, but none with ray-structure;

2 = auroral forms with ray-structure occurred, or also flaming
aurora.

The statistics of the 367 days on which reliable observations of aurora
were possible gives the following distribution of magnetic and auroral
character-figures:

Auroral character A	Magnetic character C			Average C
	0	*1*	*2*	
0	107	43	2	0·3
1	31	104	13	0·9
2	5	39	23	1·3
Average A	0·3	1·0	1·6	

The two nights with magnetic character 2 and no aurora are not
remarkable exceptions to the general relationship; for in one case, aurora

was observed near by, and in the other case, the magnetic storm had subsided before dark. But the other extremes are worth mentioning: on five nights, during quiet magnetic conditions, ray forms of short duration were observed, not only at Lerwick, but elsewhere too.

This latter remark confirms some earlier observations by Vegard [30], who also found some exceptional but significant cases of strong moving arcs and draperies, which disturbed the magnetic field below them only slightly. In fact, though there is a marked general relation between
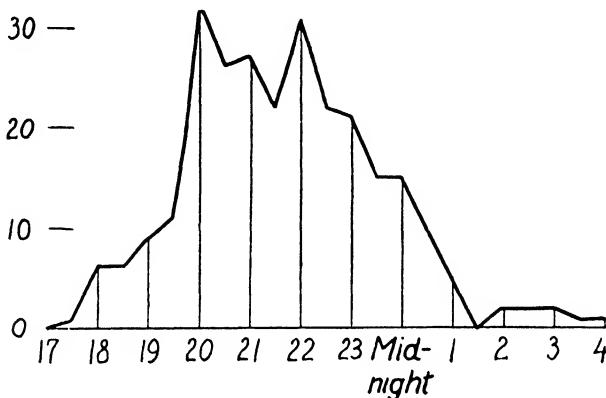
FIG. 9. The average variation of auroral frequency at Oxford (1858–72) throughout the night. (After Fritz)

aurorae and magnetic storms, it is difficult to find a detailed association between the magnetic changes and the simultaneous observations of aurorae; but usually forms with ray-structure are accompanied by more violent magnetic disturbance than quiet homogeneous forms. Rooney [44 a] finds that the relationship to auroral activity is stronger in earth-currents than in magnetic activity.

The connexion between aurorae and magnetic disturbance explains why the auroral frequency shows much the same periods as magnetic activity. The daily variations of auroral activity are, however, somewhat obscured because daylight and, to a certain degree, moonlight, outshine aurora. But the tendency for recurrence after 27 days (the solar rotation period), and the 11-year periods of sunspots, are strongly marked in the variations of auroral activity.

In the non-polar regions of Europe and North America aurora is most frequent in the earlier night hours, with a maximum at about 22^h local time (Fig. 9). This agrees well with the daily variation of magnetic activity at the same stations. At Lerwick, ray forms are most frequent

at 21^h (when they are seen on 11 per cent. of all favourable occasions), while other forms slightly increase in frequency up to the end of the observation-period, at 23^h.

In polar regions the daily variation is less uniform, and it is difficult, perhaps because of the shortness of most series, to detect common features in the observations at different stations (Hulburt [45]).

The long and consistent auroral observations on board the *Maud*
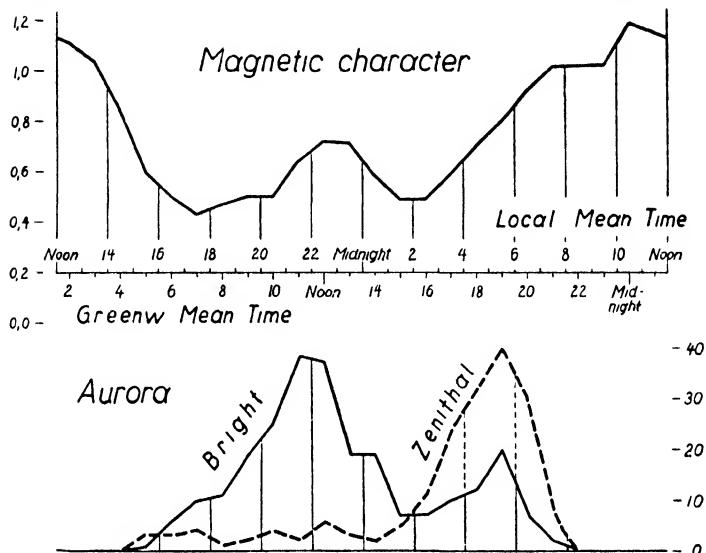
Fig. 10. The daily variation of hourly magnetic character-figure, and of auroral frequency, at Cape Denison (Australasian Antarctic Expedition, 1912–13). (After Chree)

(1922–5, to the north of Siberia), showed a maximum frequency between 22^h and 2^h, especially well-defined for auroral curtains, less definite for arches and streamers; auroral glows were distinctly more frequent near the morning. The general course of an auroral display as seen from the *Maud* is described by Sverdrup [17] as follows: 'In the later afternoon, the aurora begins with an arc and perhaps a few curtains. After 20^h the display increases in intensity until midnight, when moving forms predominate. These disappear later, leaving only a glow or an arc in the early morning hours.'

Fig. 10 shows the observations of Mawson's Australasian Antarctic Expedition, compared with the magnetic disturbance estimated by hourly character-figures (after Chree [46]). Cape Denison lies at a distance of only 14·5° from the pole of the earth's magnetic axis, well inside the southern zone of maximal auroral frequency; the aurora was,

indeed, mostly seen towards the north. The principal maximum of bright aurora, at $22\frac{1}{2}^h$ local time, nearly coincides with a secondary maximum of disturbance, and similarly for the minimum, at about $2\frac{1}{2}^h$ in the morning. The characteristic feature is the secondary maximum for bright aurora at about 6^h, which is also the hour of most frequent aurora within $20°$ of the zenith; it is possible that the decrease of auroral frequency after 6^h is simply due to the brightening of the dawn sky, without which the auroral frequency would increase right up to the principal maximum of magnetic disturbance, shortly before noon.
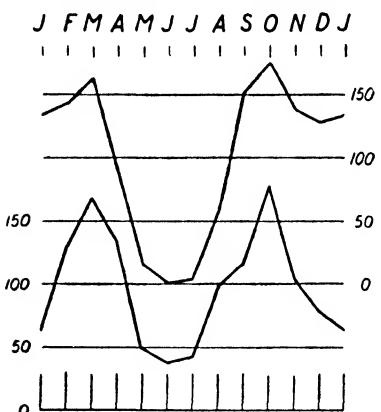
Fig. 11. The annual variation of auroral frequency, for Sweden (upper curve) and for the southern hemisphere (lower curve). The scale is adjusted so that the annual mean ordinate is 100 for each curve

The annual period of aurorae shows two peaks at the times of the equinoxes, as in the corresponding curve of magnetic activity (Fig. 11). The curve for Sweden shows the influence of the short summer nights; the curve for the southern hemisphere refers, on the average, to lower latitudes, so that the corresponding effect·(which would produce a chief minimum in December) is weaker, and in the present curve completely masked because of the deficiency of the series.

At Lerwick, the seasonal variation of auroral frequency was estimated by Lee [28] from the observations at 22^h, expressed as the number of aurorae seen during 100 favourable term-hours. In the midwinter months, December and January, this number was 32, distinctly lower than the average of 48 for the equinoctial months, September, October, March, and April.

A marked tendency of aurorae to recur after between 27 and 28 days was found by Fritz [2]. Sverdrup [17] found the same recurrence tendency in the *Maud* observations, not only for the days with unusually large auroral activity, but also for those with remarkable absence of aurora.

The 11-year cycle in the aurora australis is shown in Fig. 12, after Boller [9]. The apparent weakness of the first cycle is probably due to the scarcity of observations. The historical reports on aurora (Fritz [2, 8]) show that there were big differences in the intensity of the maxima in different 11-year periods. For instance, aurora was

visible in Rome eight times during the years 1786–9, but was rare in the next forty years. The sunspot-maxima about 1859 and 1870 brought some displays which were seen nearly all over the globe. About September 1, 1859, aurora was seen on several successive nights in Cuba and the Sandwich Islands (20° N.). The aurora of October 24, 1870, was seen at Bagdad. One of the strongest displays in history
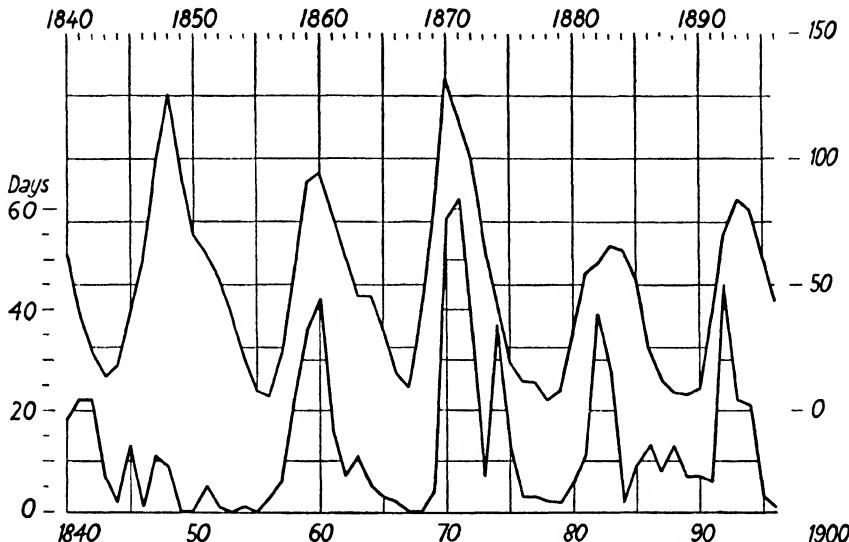
FIG. 12. The number of days per year on which southern lights (aurora australis) were seen, in 1840–96 (lower curve) compared with the relative sunspot numbers (upper curve). (After Boller)

was that of February 4, 1872. It was seen at Bombay (19° N., at 80° distance from the magnetic axis) and Aswan (Egypt) and reached the zenith at Constantinople and Athens; the simultaneous southern light reached 72° altitude at Mauritius (20° S.) and was seen by ships in the Indian Ocean between 38 and 14° S. The magnetic horizontal intensity at Bombay decreased temporarily by 750γ, about 2 per cent.

The intense magnetic storm of May 14–15, 1921, was accompanied by strong aurora. The southern lights reached 22° altitude at Samoa (13·8° S.) at the greatest intensity of the magnetic disturbance, when the horizontal intensity was nearly 3 per cent. below the normal [47]. Extensive auroral displays occurred in January and April 1938 [37, 48].

The observations of these exceptionally brilliant displays prove that the aurora may sometimes reach the zenith at places more than 50° distant from the magnetic axis of the earth. A special feature of aurora in low latitudes is the frequency of the red colour [48].

So far as observations are available, strong aurorae have always occurred in the northern and southern hemisphere at the same time. But the deficiency of the net of auroral observations does not enable definite conclusions to be drawn as to whether there are weaker displays which are confined to *one* auroral zone only. Proposals for synoptical observations of aurora have been made by La Cour [43 *a*].

14.10. Observations of the auroral spectrum. The history of the measurement and interpretation of the auroral spectrum is marked by some important discoveries in geophysics and pure physics.

Rainbows and haloes, which are due to reflected or refracted light, show strong polarization. The possibility that the aurora was of similar origin had to be discarded when no trace of polarization was found in its light [1, p. 41]; see also Harang [59]. The aurora is therefore self-luminous. In 1838 Gauss [49] remarked that in the aurora 'there is every appearance that electricity in motion performs a principal part'; Dalton [6 *a*] and perhaps others had previously made similar conjectures.

Ångström was the first, in 1867, to observe the spectrum of the aurora; he found a strong line in the yellow-green about 5,570 (Ångström units, 10^{-8} cm.), three other lines towards the blue, and in reddish aurorae a red line at about 6,300 A. Other observers found many more lines, but the spectrographs used had such small dispersion, that is, the length of the total spectrum on the plate was so small, that the wave-lengths were very uncertain; it was even difficult to find out whether different observers referred to the same line. Kayser, in his *Handbuch der Spektroskopie*, vol. 5 (1910), characterized the measurements as very deficient, and inadequate as a basis for any conclusions as to the chemical origin of the spectral lines.

Early progress was due to Vegard [3], who on his expedition to Bossekop, 1912–13, and later, used spectrographs of high light-power and dispersion. The light-intensity of the aurora is so weak that it was at first hopeless to obtain spectrograms of single phases of a display; so a kind of average spectrum was obtained by exposing the same plate night after night, up to a month. Even then only the strongest lines could be detected on the plates. Thirty-nine lines were measured between 6,600 and 3,100 A; the short wave-lengths were photographed with a quartz spectrograph. Only half of these thirty-nine lines were strong and sharp enough to be measured to 1 A; owing to the variability of the auroral spectrum, several lines were obtained only on a few occasions. Other measurements were made by Lord Rayleigh [50, 51], Sommer [52–4], Slipher [55, 56], Vegard [61–3], Kaplan [65, 66], Dufay

PLATE 32
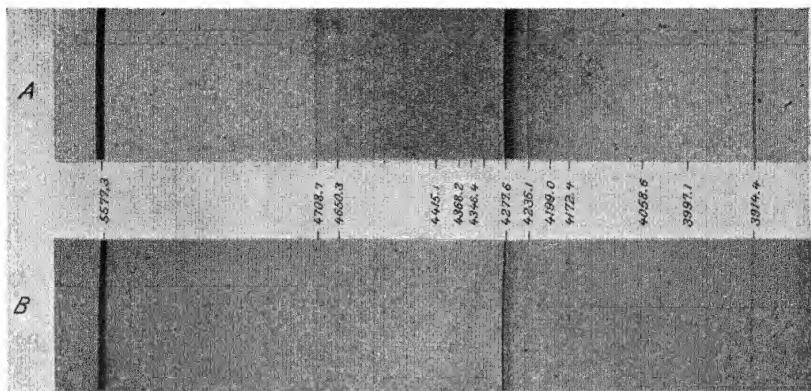
a. Two auroral spectrograms taken by Vegard and Tönsberg, A with effective exposure 37 hours (from 1935 October 15 to 1936 March 28), and B with effective exposure 15 hours (from 1936 December 3 to 1937 April 10)
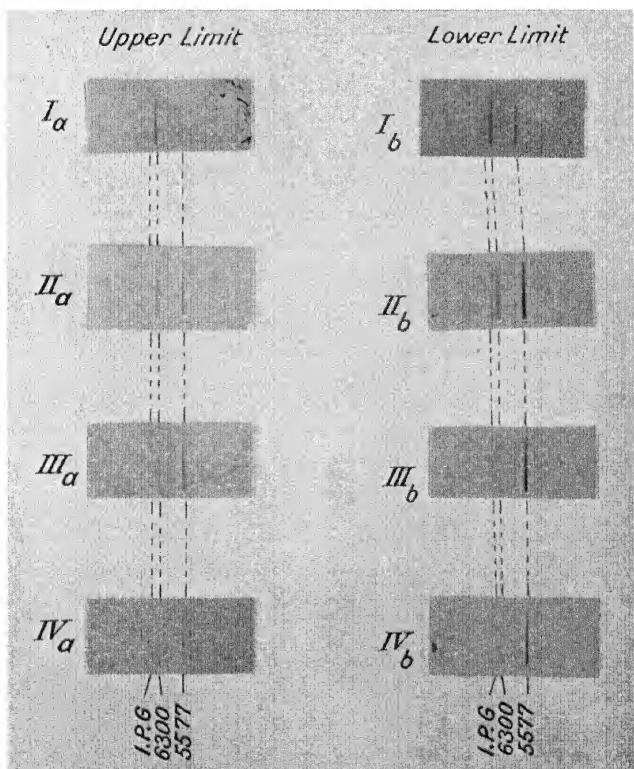
b. Spectrograms showing the enhancement of the red OI-line (6300) relative to the green OI-line (5577), with increasing altitude; I, 1937 Oct. 11, II–IV, 1937 Oct. 12 (Vegard, Geof. Publ. xii, No. 5). I.P.G. = 1st positive bands of nitrogen

PLATE 33

b. The same spectrum (on the left) compared with a spectrum of the light from yellow-green auroral curtains, height of lower border about 92 km.

a. Spectrum of the summits of blue auroral rays at a height of about 400–650 km. taken at Oslo by Störmer, 1938 September 15
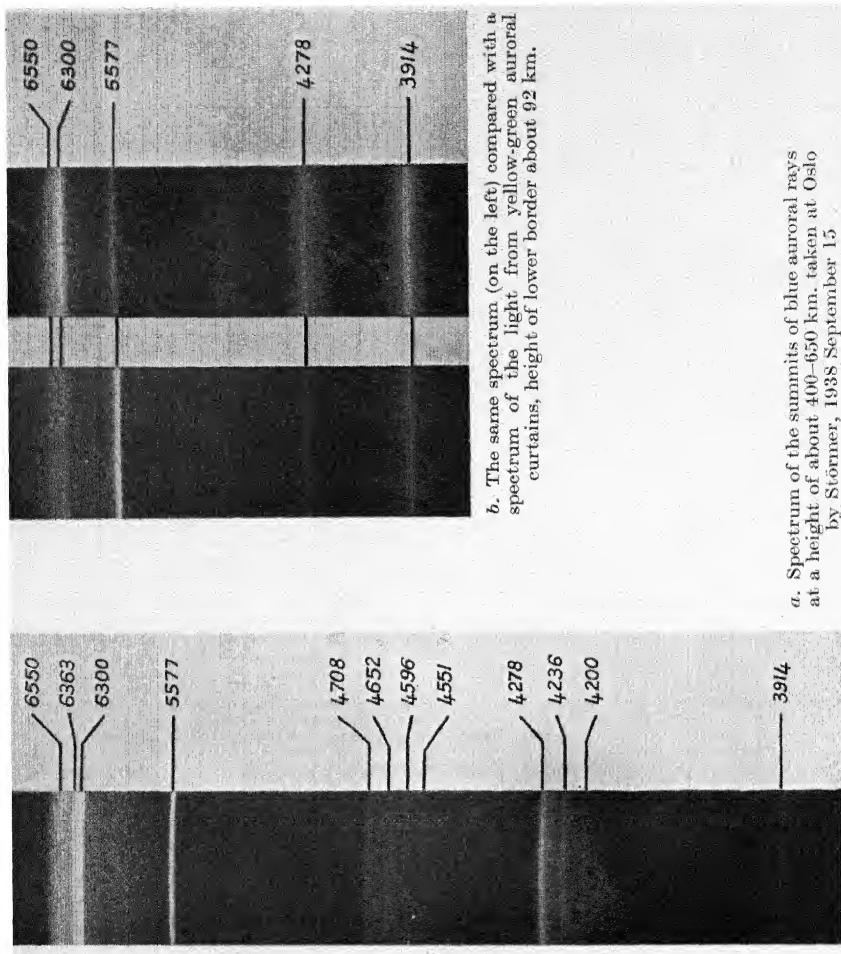
and Gauzit [64], and Störmer [36–7]. McLennan and Ireton [58] developed spectrographs which gave the three strongest lines of an aurora with exposures as short as 5 minutes; but even with an exposure of half an hour, six other bands were recorded only faintly. Infra-red spectra were obtained in 1932 by Harang and Bauer. Vegard and Harang obtained the red and infra-red spectrum between $\lambda\, 5{,}238$ and $\lambda\, 8{,}132$; they found a number of nitrogen lines, but no trace of the red lines of hydrogen or helium.

14.11. The nitrogen bands. Those parts of the auroral spectrum which were the first to be identified with spectra known from the laboratory are bands emitted by nitrogen. Singly charged molecular ions N_2^+, that is, molecules which have lost one electron, emit a group of bands called 'first negative bands' (because in a discharge tube they are emitted by the nitrogen near the negative pole, the cathode, in the negative glow, where the electric field is high and the electrons fast). Nitrogen molecules are ionized by the impact of electrons which have been accelerated by a field of 16·5 volts (the ionization potential); the head of the leading negative band $\lambda\, 3{,}914$ requires another 3·1 volts for excitation, or 19·6 volts in all. The negative bands with the heads at $\lambda\lambda\, 3{,}914$, $4{,}278$, $4{,}708$, and (less prominent) $5{,}227$ are outstanding features of the spectrum of the aurora (see Plates 32, 33). The bands are 'shaded' towards the ultra-violet (downwards in Plates $32\,a$, 33), that is, the frequencies of their lines are higher than those of their heads. These bands in the violet and blue give the most actinic light of the aurora.

The band-spectrum of neutral nitrogen molecules appears in the auroral spectrum in a number of bands of low intensity which belong to the 'second positive band system'. They are called positive, because they are emitted near the anode of a discharge tube by the so-called positive column, where the electrons are too slow to produce ions. On Plate 32 the heads of two of these bands, at $3{,}997$ and $4{,}059$, are faintly visible.

The energy levels of nitrogen and oxygen, with a discussion of the various band-spectra produced in the atmosphere, are given by Saha [15.41 a].

Vegard [3] investigated the spectrum of the light emitted by solid nitrogen under bombardment by electrons, and used his results in an attempt (given up later) to explain thus the entire auroral spectrum, inclusive of the green line. His hypothesis that the auroral spectrum was due to the luminescence of solid small nitrogen crystals in the high

atmosphere encountered the objection that the necessary low tempera-
ture of $-210°$ C. was highly improbable, and the hypothesis lost its
main attractive feature when the real nature of the green line was
discovered.

**14.12. The green line and other atomic lines of oxygen and
nitrogen.** The most important line in the auroral spectrum is that
near $\lambda 5,577$. Its origin was for a long time a puzzle to the spectro-
scopists, and the history of its final
identification as a line of atomic
oxygen is remarkable from both the
physical and geophysical aspects.

Wiechert [67] found in 1902, at
Göttingen, that this line can be seen
in the spectrum of the night sky on
every clear night, whether aurora was
indicated in the usual forms or not;
even moonlight does not outshine it.
Slipher [56] obtained photographs of
this line on every spectrogram of the
night sky. With a powerful three-prism
spectrograph, and exposures as long
as 100 hours, he obtained the wave-
length $\lambda 5,578$; the values found by
others (for actual aurorae) differed by
several units among themselves and
from the value measured by Slipher;
this was due to the insufficient accu-
racy of the older observations.

Fɪɢ. 13. To illustrate the interferometric
measurement of the wave-length of the
green line λ 5,577 in the night-sky light
or aurora
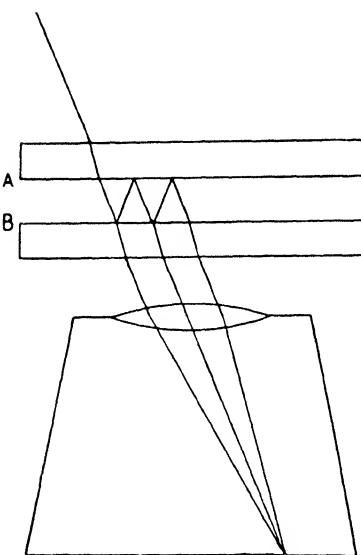
Lord Rayleigh [50] was able to photograph the green line in two out
of every three nights in England. He called this light the 'non-polar
aurora', because it occurs all over the earth; a further distinction from
the displays of ordinary 'polar aurora' is the comparative weakness
or absence of nitrogen bands in its spectrum (see § 13).

Babcock [68] measured the wave-length of the green line very accu-
rately by photographing the night sky for several hours, with an
ordinary camera, before which a Fabry-Pérot interferometer had been
placed (Fig. 13). The latter consists of two plates of plane-parallel glass.
The surfaces A and B are covered with very thin metallic films; Babcock
used gold cathodically deposited, because this metal transmits green
light freely. Parallel light, transmitted either directly, or after single

PLATE 34
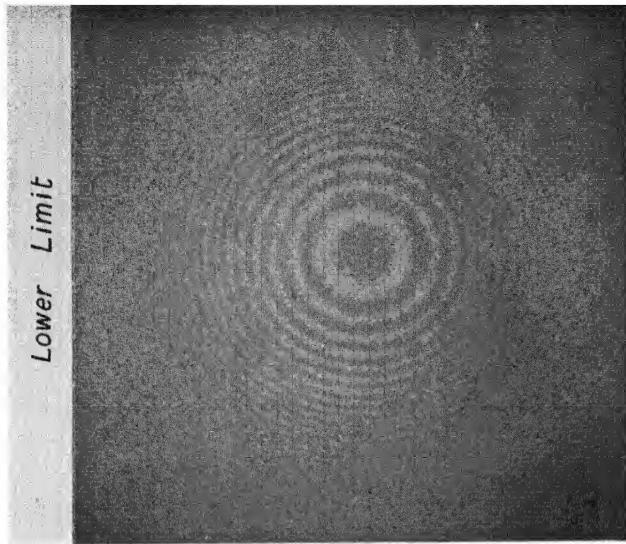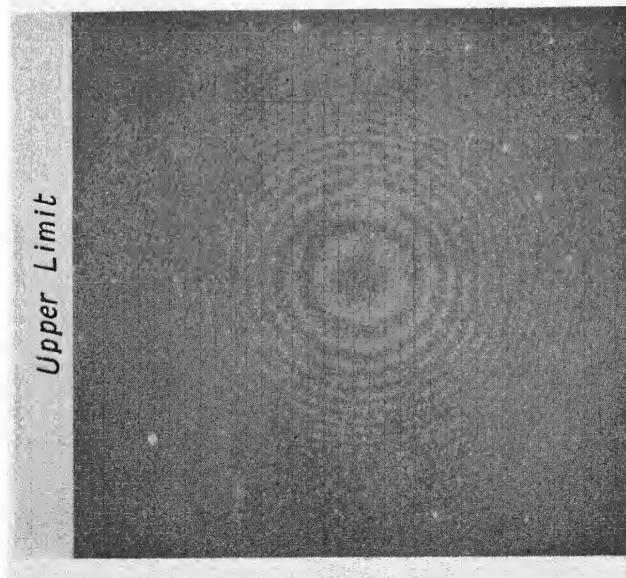
Interferometric photographs by Vegard, in the light of the green auroral line λ 5,577, from two levels, 100 km. apart, of the aurora of 1937 October 12–13

or multiple reflection at the opposed surfaces, is focused on one and the same point of the photographic plate. A singly reflected beam will have an optical path which exceeds that of the beam directly transmitted by somewhat more than double the distance of the plates. If the excess is equivalent to an entire multiple of wave-lengths, the direct and all the reflected beams of monochromatic light have the same phase and will produce a bright spot; otherwise, the beams will interfere more or less destructively. With a fixed distance of the plates the path-difference of the direct and the reflected beams depends only on the angle of incidence. Diffuse light of the same colour will therefore produce a system of alternately bright and dark interference-rings on the plate.

Such a system was actually obtained by Babcock at Pasadena on every night. Moderate cloudiness did not hinder the observations, nor did the light of the stars, nor of the moon at quarter phase. The wave-length was determined by comparisons with the interference-system produced by well-known lines near by, and was found to be $\lambda = 5{,}577 \cdot 35 \pm 0 \cdot 005$ (International Ångström). The line is very sharp, because even beams with path-differences of 85,000 wave-lengths, or nearly 5 cm. in air, gave well-defined interference-rings; from this degree of sharpness Babcock judged the width of the line to be less than 0·035 A. Vegard and Harang [62] later obtained similar interference-patterns with the light of λ 5,577 in the polar aurora (Plate 34), and found that the wave-length agreed with Babcock's measurement for the green line in the spectrum of the non-polar aurora.

In 1925 McLennan and Shrum [57] obtained the green line in the laboratory from atomic oxygen by suitable electrical excitation. The reason why this line had never been found before is its susceptibility to the conditions of the discharge. A vacuum tube discharge in pure oxygen shows the line very faintly, if at all, and masked by the band spectrum of molecular oxygen. But if a rare gas, preferably argon, is mixed with the oxygen, the green line is greatly enhanced in proportion to the other spectrum of oxygen; finally [58], the spectrum of the non-polar aurora was reproduced in a discharge in argon with only a trace of oxygen. The effect of the inert gases is perhaps due to the uniform small potential drop throughout the discharge, which limits the oxygen spectral lines to those of low excitation, such as the green line.

Interference-measurements by McLennan, analogous to those of Babcock, established definitely the identity of the oxygen line with the green line in the non-polar and polar auroral spectrum, the wave-length

found by McLennan being $5,577 \cdot 341 \pm 0 \cdot 004$, and the spectral width of the line $0 \cdot 030$ A.

In order to classify the green line, its change in a magnetic field was observed. In the language of atomic physics, the line was found to give the Zeeman pattern of a singlet line, and was therefore ascribed to the forbidden transition between the low metastable terms $2p\,^1D_2$ and $2p\,^1S_0$ in the arc spectrum of the (neutral) oxygen atom. The excitation potential necessary to bring the oxygen atom from its normal state into the higher level $(2p\,^1S_0)$ is $4 \cdot 2$ volts; the energy required to dissociate the oxygen molecule is $5 \cdot 1$ volts.

The explanation of the green line is therefore the same as for the famous lines in nebulae, which were identified by Bowen [77] as due to forbidden transitions in the nitrogen and oxygen systems of energy levels; such lines had never been observed in the laboratory. For physical considerations regarding the probable conditions of emission we refer to special papers [57]. The long time-interval between successive atomic or molecular encounters in the rarefied gases of the high atmosphere favours the emission of these spectral lines because they allow the metastable states to finish their lifetimes, which are to be reckoned in tenths of a second, by a spontaneous transition to a lower energy level.

The successful classification of the green line led to the prediction of three oxygen auroral lines in the red, at $\lambda\,6{,}300$, $6{,}364$, and $6{,}392$. This 'red oxygen triplet' was found by Sommer [54] and others in the spectrum of the polar aurora, and by Paschen [69] in the laboratory. Vegard, by interferometer measurements, found the exact wave-length of the strongest red auroral line to be $\lambda\,6{,}300 \cdot 286$ A, agreeing with the red oxygen line O 1 $(^1D_2 - {}^3P_2)$ for which Hopfield, in the laboratory, had found the wave-length $\lambda\,6{,}300 \cdot 279$. This and the next strongest line of the oxygen triplet are visible in the spectrograms of Plate 33.

Slipher and Sommer [55] found in the auroral spectrum a line at $\lambda\,5{,}206$, which they ascribed to a transition between low metastable terms in the neutral nitrogen atom; it is excited by electrons of $2 \cdot 4$ volts energy. This line, as well as the green line $5{,}577$, is also present as an absorption line (due to solar nitrogen) in the solar spectrum [53]. Vegard and others have also found lines due to ionized atoms of oxygen and nitrogen; for example, Vegard ascribes the line $4{,}368$, shown in Plate 32 a, to ionized atomic oxygen, but expresses doubt whether any of the lines can be ascribed to neutral atomic nitrogen.

Thus the spectroscopic analysis of the aurora has indicated the

presence, in the upper atmosphere, of neutral and singly ionized mole-
cules of nitrogen, and of neutral and ionized oxygen and nitrogen atoms.
Several lines observed in the polar aurora have not yet been definitely
interpreted.

14.13. The non-polar aurora. The light of the night sky has now
been extensively observed; the results have been summarized by
Déjardin [70] and by Chapman and Price [72, 15.1]. Lord Rayleigh [50]
found two other lines at $\lambda 4,419$ and $\lambda 4,168$, but never nitrogen bands
except when actual polar aurora was indicated by one of its charac-
teristic forms in the sky. Slipher [56] has, indeed, photographed the
nitrogen bands in the west just before the last glimmer of daylight dis-
appeared; but Gauzit [15.19 c,] states that neither he, nor Cabannes
and Dufay, nor Cario, have been able to detect them, and Sommer's
[52] observations of them at Göttingen may possibly have been taken
during actual polar aurora. The recent work of Cabannes, Dufay, and
Gauzit has yielded a great number of night-sky lines, some of which
could be identified with nitrogen bands, as in the auroral spectrum,
though they were weaker than there, compared with the green line.

The intensity of the non-polar aurora has a maximum near midnight
(McLennan [57] and Rayleigh [50]). Photometric measurements with
the use of colour filters, which divide the spectrum into three regions,
red, auroral, and blue, were continued for some years at several stations
in different latitudes [50]. The intensities were found to vary con-
siderably from night to night, the maximum being about three to four
times the minimum; the red and auroral parts of the light vary nearly,
but not quite, in proportion. Definite correlations between the values
obtained at different stations or with magnetic activity have not yet
been found.

An average value for the absolute intensity of the non-polar night-
sky light was estimated by Lord Rayleigh [51]. The brightness of the
sky produced by the green oxygen line is $3 \cdot 2 \times 10^{-5}$ candles per square
metre; the energy required to maintain it is $6 \cdot 4$ ergs per second per
square metre, and the number of atomic transitions required to supply
this energy is $1 \cdot 8 \times 10^8$ per second in a vertical column of one square
centimetre horizontal cross-section throughout the atmosphere.

Déjardin [70], summarizing the excellent work done by the French
physicists Cabannes, Dufay, and Gauzit, characterizes the light of the
night sky as follows:

(*a*) The brightness of the clear night sky, when the sun and moon
are 18° or more below the horizon, produces an illumination of the

horizontal plane of 3×10^{-4} lux, equivalent to that produced by a 25 candle-power lamp at a distance of about 300 metres. Only 20 to 25 per cent. of this illumination is caused by starlight; the rest is produced by an emission of light in the high atmosphere.

(b) Dufay found the light of the night sky to be weakly polarized with the plane of polarization passing through the sun, but the proportion of polarized light is only 2 to 4 per cent. of the total.

(c) The green line $\lambda 5,577$ produces between 6 and 9 per cent. of the total brightness of the night sky (Dufay, Lord Rayleigh).

(d) The essential difference between the bands in the night-sky spectrum and in the auroral spectrum is in the degree of excitation. In the night sky the Vegard-Kaplan nitrogen bands between 3,500 and 5,000 A (energy 6·1 volts) are the most intense, then come the bands of the first and second positive systems (7·4 and 11·0 volts); lastly, the negative bands (19·6 volts) have a much reduced intensity. The order of intensity is exactly reversed in the auroral spectrum.

(e) According to Lord Rayleigh and Spencer Jones [74] observations made at Terling (England), Capetown, and Canberra (Australia) between 1925 and 1933 in three regions of the night-sky spectrum, viz. around the green auroral line, in the blue, and in the red, show that the brightness near $\lambda 5,577$ varies very much with the solar cycle; there is also a slight indication of greater brightness in this region on magnetically disturbed days.

14.14. Variations in the auroral spectrum. The differences between the spectra of different aurorae, and at different heights of the same aurorae, have been studied especially by Vegard and Störmer. In view of the great range of auroral colours, considerable differences in the relative intensities of the various spectral constituents of auroral light are to be expected, and are actually found. Some of Vegard's conclusions on this subject are as follows [63 b]. (i) The intensity of the red oxygen triplet increases very much with height, relative to that of the oxygen green line. (ii) The relative intensity of the first positive group of molecular nitrogen bands rapidly diminishes upwards. (iii) The red fringe sometimes observed at the lower border of aurorae is ascribed to the occasional great intensity of these first positive bands at the lower auroral levels.

Plate 33 shows two recent auroral spectra obtained by Störmer. The upper picture shows the spectrum of the light from the summits of blue sunlit auroral rays at a height of 400 to 650 km., taken at Oslo on 1938 September 15. The lower picture shows this spectrum (on a reduced

scale of wave-length) set alongside another spectrum of a lower aurora, consisting of yellow-green curtains in the earth's shadow; both spectra were taken with the same spectrograph on similar photographic plates. The great strength of the yellow-green line $\lambda 5,577$ in this second spectrum is evident. The upper spectrum shows the relative intensification of the red oxygen triplet at $\lambda 6,300$, $\lambda 6,363$, $\lambda 6,398$ at the high levels, and also of the blue bands at 4,278 and 3,914, which give the blue colour to this aurora; the latter are bands of the negative nitrogen group. The weaker members of this group, namely 4,236 and 4,200, are also stronger, compared with 4,278, than in ordinary aurorae, and similarly for the bands 4,551–4,652, relative to 4,278.

14.15. The air-temperature at auroral levels. Vegard has made important studies on the temperature of the air at auroral levels, as determined from the intensity distribution among the lines of the rotational spectral bands of nitrogen in the auroral spectrum. Plate 32 shows the spectra used by Vegard and Tönsberg for this purpose. They found strong evidence, in this way, indicating that the air-temperature between levels 110 and 150 km. above the ground is $-30°$ to $-40°$ C. (or about 240° absolute). Vegard also made interferometer measurements on the breadth of the green and red lines of atomic oxygen; Babcock had already found that the line $\lambda 5,577$ in the spectrum of the night sky is very sharp and narrow, a fact which enables an upper limit to be set to the possible temperature of the air at the level of emission, because the random velocities of the emitting particles modify the wave-length of the received light, owing to the Doppler effect. Vegard found that the breadth of the lines of oxygen in the spectrum of the polar aurora is likewise very small, and shows no appreciable change with height, between 100 and 300 km. (Plate 34 shows this for a smaller height-difference, of 100 km.)

14.16. Relations between cosmic rays and terrestrial magnetism. Cosmic rays were discovered by Hess in 1912 by their ionizing effect on the air in closed vessels taken up in balloon flights; this effect increases upward, and is present by day and night without conspicuous changes. The intensity of cosmic radiation is measured in the unit J, equal to the production of 1 ion-pair per cm.^{3} per sec. in air at atmospheric pressure. Cosmic rays are fast particles, which have energies equivalent to those acquired by electrons falling through electric fields of from about 10^9 volts to over 10^{11} volts; they come from all directions of space and penetrate the atmosphere from outside; they lose their energy by producing 'showers', consisting of a large number of secondary
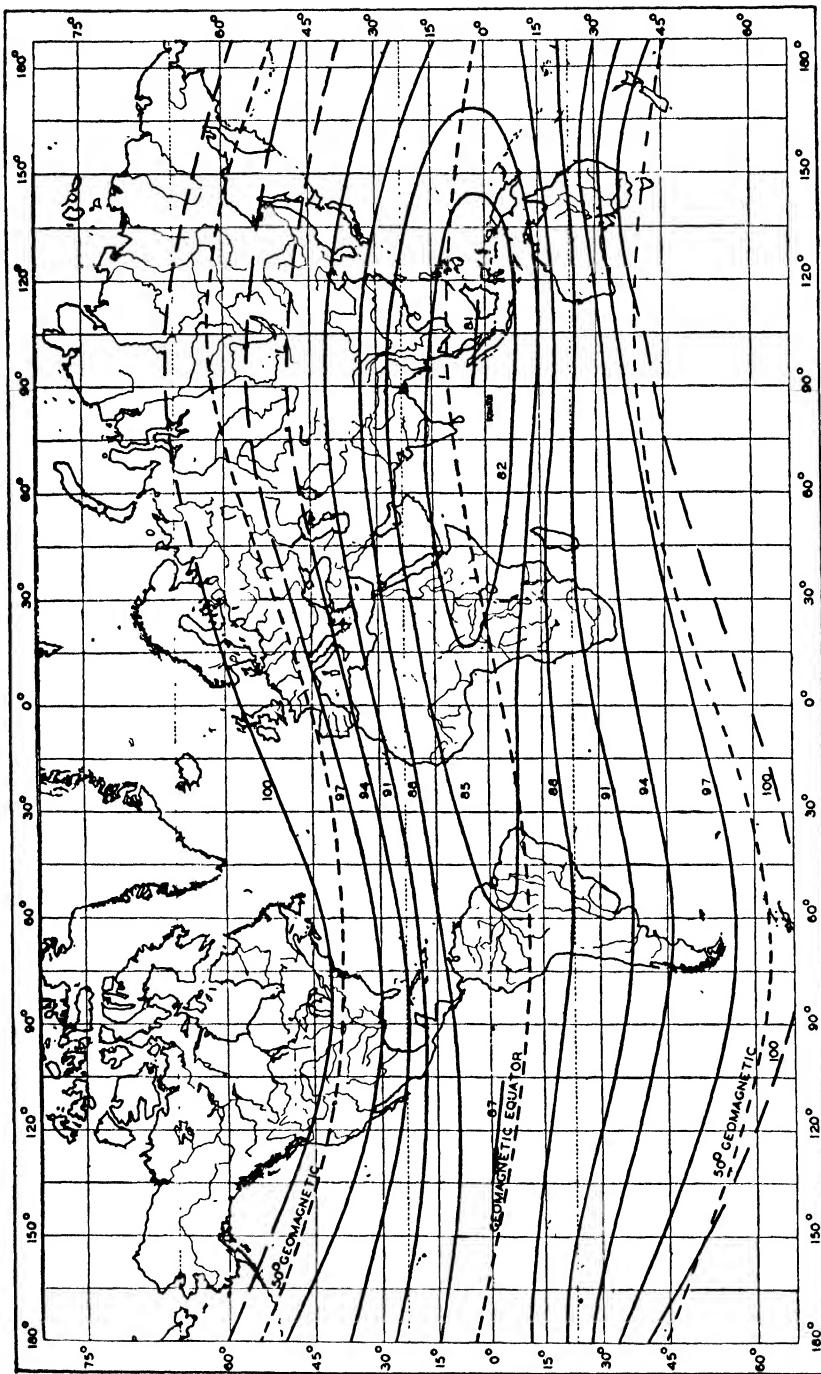
Fig. 14. Lines of equal cosmic-ray intensity. (After A. H. Compton)

electrons, positrons, and mesons.   At least 98 or 99 per cent. of these particles are electrically charged.

Charged particles approaching the earth from outside are subject to a deflexion by the earth's magnetic field, much as the solar particles producing aurora (Ch. XXIV). They are therefore more abundant in polar regions, although their great energy allows them to penetrate to the ground even at the earth's equator. At sea-level, the intensity increases by about 16 per cent., from about 1·46 J at the equator, to
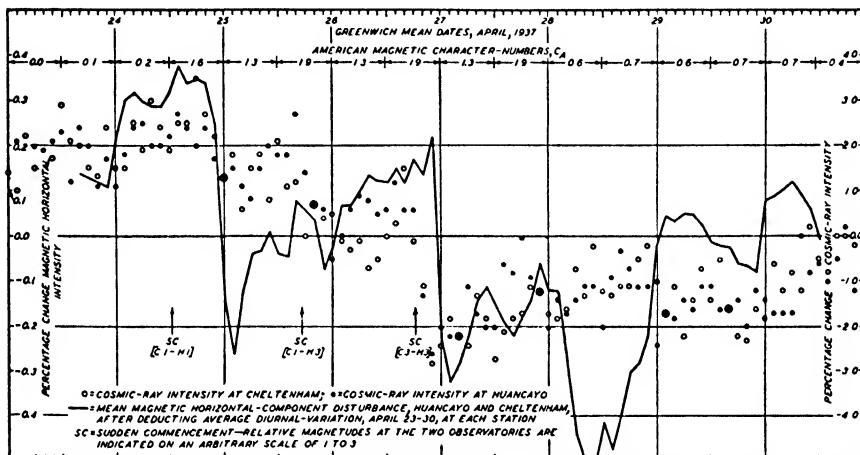
FIG. 15. Simultaneous variations of cosmic-ray intensity and of horizontal magnetic force, at each of two stations (Cheltenham, Maryland, U.S.A. and Huancayo, Peru), on 1937 April 24–30. (After Forbush)

about 1·70 J at 50° geomagnetic latitude; it then remains constant to the pole (cf. Fig. 14, where the polar value is put equal to 100). At high altitudes the increase is greater; at about 15 km. height the intensity increases in the ratio 1 : 25, from about 12 J over the equator, to 300 J over 50° geomagnetic latitude. The equatorial asymmetry (18.6) of the earth's magnetic field—low values of H over the Atlantic, high values over Siam—seems also to be indicated in cosmic-ray intensities (in the reverse direction, low values of ionization over the East Indies), although the evidence is perhaps not yet quite conclusive.

Near the equator, and at high altitudes, more particles (presumably positive) arrive from the west than from the east; uni-directional recording will further analyse the latitude-effect.

Cosmic radiation—which is being intensely studied [78–84]—is connected with geomagnetism by the fact, already mentioned, that its geographical distribution appears to be related to the geomagnetic latitude.

Recently, a direct influence of a magnetic storm on the intensity of cosmic radiation near sea-level all over the globe has been found by Forbush [86] (Fig. 15). At Cheltenham (Maryland) and Huancayo (Peru), the Compton-Bennett cosmic-ray meters showed a striking decrease, by more than 4 per cent., from 1937 April 24 to April 27, and Hess [87] found the same amount for a station near Innsbruck (Austria) at 2,300 m. height. The horizontal intensity during the magnetic storm decreased by about 0·4 per cent. during the same interval. Similar reductions of cosmic-ray intensity during the great magnetic storms of January and April 1938 have been reported [90] and discussed [88–9]; while the phenomenon itself is established beyond doubt, its physical nature is not yet clearly understood.

XV

THE EARTH'S ATMOSPHERE

15.1. Troposphere and stratosphere. The present chapter gives a brief account of various properties of the earth's atmosphere which are shown, in Part 3 of this book, to bear on the physical theories of some of the geomagnetic phenomena described in the preceding chapters.

The atmospheric temperature in general decreases upwards from the ground, to a height which varies with the latitude, being about 16 km. at the equator and about 9 km. at the poles. This layer of upward-decreasing temperature is called the *troposphere*. Above it is a layer called the *stratosphere*, in which, for many kilometres upwards, the temperature remains almost constant in the vertical direction.

The average temperature of the tropospheric air near the ground varies from about 26° C. at the equator to about −24° C. near the poles; there are, of course, annual and daily variations of air temperature near the ground, which on land are considerable. The average rate of decrease of temperature with height does not vary much with the latitude; it is about 6° C. per km. Since the troposphere is much thinner at the poles than at the equator, the temperature at the top (which is also that of the lower stratosphere) is considerably lower over the equator than over the poles, despite the higher ground temperature at the equator; the values are approximately −80° C. at the equator and −50° C. at the poles.

Though air-temperatures are often expressed, as here, on the Centigrade scale, for our purpose it is more convenient to use the absolute or Kelvin scale (0° C. = 273·1° K.); this will be done throughout the rest of this chapter. The temperature so expressed will be denoted by T. Thus the average value of T at points in the troposphere ranges from about 299° to 193° (respectively near the ground and at the top of the troposphere, above the equator), though more extreme values are registered at times, ranging from 330° at the ground to 182° at the top of the troposphere.

The ordinary weather phenomena of clouds and rain are located in the troposphere. The continual winds and turbulence in this region maintain the air-composition practically uniform throughout it at all times and places, so far as concerns the permanent constituents (namely those which do not undergo appreciable physical or chemical change

in the troposphere). The water vapour content, of course, varies considerably, owing to evaporation, condensation, and precipitation. The composition of dry tropospheric air, per cent. by volume, is approximately as follows [6]; the small percentages of carbon dioxide, radon, and ozone, however, are subject to variations which, though small absolutely, are proportionately considerable.

TABLE 1. *Composition of dry tropospheric air (per cent. by volume).*

Molecular nitrogen	78·08 ⎫
„ oxygen	20·95 ⎪ 99·99
Argon	0·93 ⎬
Carbon dioxide	0·03 ⎭
Neon	$1·8 \times 10^{-3}$
Helium	5×10^{-4}
Krypton	1×10^{-4}
Xenon	1×10^{-5}
Ozone	variable; $> 1 \times 10^{-6}$
Radon (average near ground) . .	6×10^{-18}
Hydrogen	5×10^{-5}

The atmospheric properties of geomagnetic interest are mainly those of the upper layers. Important *electrical* phenomena, especially those connected with thunderstorms, occur in the troposphere, but their magnetic effects are small, local, and fleeting. The electrical conductivity and current-intensity in the atmosphere seem first to become important (for geomagnetism) at levels well above 50 km.

At any level of the atmosphere the primary data are the pressure, temperature, and composition, which are inter-related as described in § 2. These quantities are known by direct observation to a height of rather more than 20 km. Our knowledge of the properties above that level is extensive, but in many ways incomplete and uncertain.

15.2. Pressure, density, and temperature [1]. Vertical accelerations in the atmosphere are very small except for brief periods and in limited regions: hence the height-distribution of the atmospheric pressure p is effectively governed by the statical equation

$$\frac{dp}{dh} = -\rho g, \tag{1}$$

where h denotes height, ρ the air density, and g the acceleration of gravity. In c.g.s. units, p is expressed in dynes/cm.^{2} Meteorologists call a pressure of 1 dyne/cm.^{2} a microbar, and usually take as their unit of pressure the millibar, equal to 1,000 dynes/cm.^{2}; these names are associated with the fact that 10^6 dynes/cm.^{2}, or 1 bar, is approximately equal to the normal atmospheric pressure; 1 bar is in fact equivalent

to the pressure of a column of mercury 750·1 mm. or 29·531 inches high at 0° C. in latitude 45°.

If the molecular masses of the various atmospheric constituents are m_1, m_2,..., and if at height h there are n_1, n_2,... molecules of each, per unit volume (1 c.c.), then

$$\rho = n_1 m_1 + n_2 m_2 + ... \tag{2}$$

(It is convenient to refer to n_1, n_2,... as number-densities.) Also, if k denotes Boltzmann's constant ($= 1·372 \times 10^{-16}$ in c.g.s. units),

$$p = k(n_1 + n_2 + ...)T, \tag{3}$$

ignoring the small corrections included in such formulae as that of van der Waals. It is convenient to write

$$n_0 = n_1 + n_2 + ..., \qquad n_0 m_0 = n_1 m_1 + n_2 m_2 + ... \tag{4}$$

so that n_0 is the total number-density of molecules (including independent atoms and electrons and ions), and m_0 is the mean molecular mass. Then (2) and (3) may be written in the equivalent forms

$$\rho = n_0 m_0, \tag{5}$$

$$p = k n_0 T, \tag{6}$$

and (1) may be rewritten as

$$\frac{1}{p}\frac{dp}{dh} = -\frac{\rho g}{p} = -\frac{n_0 m_0 g}{k n_0 T} = -\frac{m_0 g}{kT} = -\frac{1}{H_0} \tag{7}$$

or

$$\frac{d\log_e p}{dh} = -\frac{1}{H_0}, \tag{8}$$

where $\log_e$ denotes the Napierian logarithm, and H_0 is a quantity, having the dimensions of a length, defined by

$$H_0 = kT/m_0 g. \tag{9}$$

Again, if we write

$$p_1 = k n_1 T, \qquad p_2 = k n_2 T, \quad ..., \tag{10}$$

$$\rho_1 = n_1 m_1, \qquad \rho_2 = n_2 m_2, \quad ..., \tag{11}$$

p_1, p_2,... denote the 'partial pressures', and ρ_1, ρ_2,... the 'partial densities' of the various constituents: clearly

$$p = p_1 + p_2 + ..., \qquad \rho = \rho_1 + \rho_2 + ... \tag{12}$$

If the constituent 1 were alone present, and subject only to gravity, p would equal p_1, and (8) would become

$$\frac{d\log_e p_1}{dh} = -\frac{1}{H_1}, \tag{13}$$

where

$$H_1 = kT/m_1 g. \tag{14}$$

The quantity H_0 defined in (9) is related to H_1, H_2,... for the several constituents separately, by the relation

$$\frac{1}{H_0} = \frac{n_1/n_0}{H_1} + \frac{n_2/n_0}{H_2} + ...; \qquad (15)$$

that is to say, H_0 is the *weighted harmonic mean* of the constants H_1, H_2,..., the weights being the ratios n_1/n_0, n_2/n_0,... which are the proportions by *volume* (or *concentration ratios*) of the respective constituents.

In an atmosphere which is at uniform temperature T throughout, and which is also uniform in composition, so that m_0 is constant, the quantity H_0 is also constant, at least over the range of height in which g can be regarded as constant (actually g decreases inversely as the square of the distance from the centre of the earth, or by about 1 per cent. for a height increase of 32 km. at moderate heights). The equation (8) then has the simple solution

$$p = p_0 e^{-h/H_0}, \qquad (16)$$

that is, the pressure decreases upwards exponentially; p_0 here denotes the pressure at ground-level, where $h = 0$. In these circumstances, also, $\rho \propto p$, and

$$\rho = \rho_0 e^{-h/H_0}, \qquad (17)$$

where ρ_0 is the density at ground-level. The total mass (M) of the atmosphere, per vertical column of unit cross-section, is then $\int_0^\infty \rho \, dh$ or $\rho_0 H_0$, the same as for an atmosphere of *uniform* density ρ_0 and total height H_0. For this reason, and in reference to these special conditions, H_0 is often called the 'height of the homogeneous atmosphere'.

In an atmosphere wholly composed of constituent 1, at uniform temperature, $p = p_1$, $H_0 = H_1$, and (16) and (17) indicate that p_1 and ρ_1 vary as e^{-h/H_1}.

In the earth's atmosphere H_0 varies with the height, and the simple solution (16) of (8) is not valid. Nevertheless H_0 remains of importance as a measure of the rate of *proportional* variation of the pressure with height: this may be illustrated by rewriting (8) in the form

$$\frac{d \log_e p}{d(h/H)} = -1, \qquad (18)$$

appropriate to any level h; here H is to be regarded as a constant with the value of H_0 at the level h; H_0 may then be interpreted as a unit of height-measurement relative to which, at the given level, the rate of upward decrease of $\log_e p$ is unity.

It is convenient to adopt a name for H_0 in general; the name 'height

of the homogeneous atmosphere' is clearly not appropriate when H_0 varies with height; the name (*local*) *scale-height* will be adopted here. A similar significance can be given to the quantities H_1, H_2,... in pure atmospheres of one constituent only, but not necessarily at uniform temperature. In reference to a mixed atmosphere, they may be called the *partial* scale-heights. The relation (15) connects the *mean* scale-height with the partial scale-heights at any level. It is hardly necessary to consider the case when the different constituents of the atmosphere have different temperatures T_1, T_2,... in the same region, though the above equations can readily be modified to take this into account if necessary. Ordinarily the continual interchange of molecular energy at collisions maintains T_1, T_2,... very nearly equal; an appreciable inequality between such partial temperatures might, however, exist in a region where ionization or dissociation is proceeding at a rapid rate, if the products of the process are released with translatory energies much in excess of those corresponding to the local mean temperature.

The logarithms in the above equations are the natural or Napierian logarithms. Since

$$\log_e x = 2 \cdot 303 \log_{10} x, \tag{19}$$

where $\log_{10}$ denotes the common logarithm, to the base 10, (18) may be written

$$\frac{d \log_{10} p}{d(h/H')} = -1, \tag{20}$$

where

$$H' = H'_0 = 2 \cdot 303 H_0 = 2 \cdot 303 \, kT/m_0 g. \tag{21}$$

Over an interval in which H_0 is constant, (16) and (17) may in like manner be modified to

$$p \propto 10^{-h/H'}, \quad \rho \propto 10^{-h/H'}, \tag{22}$$

so that *if h is increased by H', p and ρ are reduced in the ratio $\frac{1}{10}$*. If H' is not strictly constant, but varies only slightly over a range of h of amount H', the ratio of reduction of p and ρ over this range will still be approximately $\frac{1}{10}$. On this account it is convenient to call H' the *decimal* (local) scale-height of the atmosphere; similarly H'_1, H'_2,..., equal to $2 \cdot 303$ times H_1, H_2,..., are the decimal partial scale-heights. Clearly (cf. (15))

$$\frac{1}{H'_0} = \frac{n_1/n_0}{H'_1} + \frac{n_2/n_0}{H'_2} + \dots. \tag{23}$$

It is of interest to tabulate H' for some of the more interesting known or possible atmospheric constituents. Since the values of H' depend on T, it is convenient to express them in terms of their special values, which we shall denote by H', appropriate to the special temperature

$0°$ C. or $T = T_0 = 273°$. These values may be called the *normal* decimal scale-heights. Clearly

$$H_1' = 2\cdot303\, kT_0/m_1 g, \qquad (24)$$

and

$$H_1' = (T/273)H_1'. \qquad (25)$$

In calculating H_1', the value of g has been taken as 981 cm./sec.^{2}

TABLE 2. *Normal decimal scale-heights* H' (*in km.*)
(*for* $T = 273°$ *abs.*).

Molecular nitrogen	19·0
„ oxygen	16·6
Argon	13·3
Carbon dioxide	12·0
Water vapour	29·4
Neon	26·3
Helium	133
Molecular hydrogen	264
Ozone	11·1
Atomic oxygen	33·2
„ hydrogen	528

In the troposphere the volumetric proportions n_1/n_0, n_2/n_0,... of the main atmospheric constituents, given in Table 1, are constant; they lead to the following value of the normal decimal scale-height H_0' for tropospheric air:

$$H_0' = 18\cdot4 \text{ km.}$$

The local decimal scale-height is slightly greater than this near the ground, in temperate and low latitudes, since T there exceeds 273°. At the top of the troposphere, on the contrary, where $T < 273°$, $H_0' < H_0'$; e.g. if $T = 220°$, $H_0' = 14\cdot8$ km., whereas at the equator, where $T = 193°$ in the lower stratosphere, $H_0' = 13\cdot0$ km. These lower stratospheric values of H_0' remain applicable throughout a thickness of several kilometres, but above about 30 km. the value of H_0' becomes increasingly uncertain.

If H_0' or H_0 were known as a function of h, from the ground upwards, the height-distribution of the pressure p could be determined by solving (8), whether analytically, graphically, or numerically: thus

$$\log_e p_0 - \log_e p = \int_0^h \frac{dh}{H_0}, \qquad (26)$$

or

$$\log_{10} p_0 - \log_{10} p = \int_0^h \frac{dh}{H_0'}. \qquad (27)$$

Conversely a knowledge of p as a function of h would suffice to deter-

mine H_0 at all heights, which would indicate the distribution of m_0/T, since k and g are known.

If both p and T were known as functions of the height, H_0 would be calculable from p, yielding the value of m_0/T, whence, knowing T, we can find m_0; also from (6), knowing p and T, we can derive n_0, and so can calculate the density ρ at all heights.

At the ground, in typical calculations, we may take

$$\left.\begin{aligned}
&p = p_0 = 1 \times 10^6 \text{ microbars}, \\
&n_0 = 2 \cdot 66 \times 10^{19}, \quad T = 273 \cdot 2°, \\
&\rho_0 = 1 \cdot 25 \times 10^{-3} \text{ gm./cm.}^3, \quad m_0 = 4 \cdot 69 \times 10^{-23} \text{ gm.}
\end{aligned}\right\} \quad (28)$$

15.3. The composition of the upper atmosphere.

In a static atmosphere at uniform temperature the various atmospheric constituents would each be distributed in accordance with equations such as (13), the partial density of constituent (1), for example, decreasing exponentially with height proportionately to e^{-h/H_1} or $10^{-h/H_1}$. Hence the lighter constituents, for which H' is larger, would decrease upwards more slowly than the heavier ones, so that their proportions by volume would steadily increase upwards.

If for any reason these conditions do not obtain in a portion of the atmosphere, thermal conduction, and relative diffusion, occur in a sense tending towards such a distribution. A different state might, however, be steadily maintained by thermal and other processes.

The rate of diffusion, in an atmosphere in which there are only two constituents (1, 2), and in which the temperature is not uniform, is given by the following equation [10 a]

$$v_1 - v_2 = -\frac{n_0^2}{n_1 n_2} D_{12} \left\{ \frac{d(n_1/n_0)}{dh} + \frac{n_1 n_2 (m_2 - m_1)}{n_0 \rho} \frac{d \log_e p}{dh} - \frac{\rho_1 \rho_2}{p\rho}(F_1 - F_2) + k_T \frac{d \log_e T}{dh} \right\}. \quad (29)$$

Here all the variables are supposed to depend only on the height h; v_1 and v_2 are the mean upward velocities of the two constituents, so that $v_1 - v_2$ is their mean relative velocity; $m_1 F_1$ and $m_2 F_2$ denote the upward external force on each molecule of the two kinds, so that if the gas is subject only to gravity, $F_1 = F_2 = -g$; D_{12} is the coefficient of diffusion, and k_T the thermal diffusion ratio. The four terms within the main bracket of (29) correspond to components of the relative velocity of diffusion due respectively to (a) a concentration gradient, (b) a pressure gradient, (c) external forces (this term vanishes when these are purely gravitational, since then $F_1 = F_2$), and (d) a temperature

gradient. These component velocities of diffusion tend as follows: (a) to make the composition uniform, (b) to increase the proportion of the heavier constituent in the regions of higher pressure, (c) to separate the constituents in the direction of the relative accelerations due to the external forces, and (d) to make the heavier molecules (if of size similar to the lighter ones) predominate in the colder regions, unless the forces between the two kinds of molecules are mainly electrostatic (as in an ionized gas), when the lighter particles tend towards the colder regions.

For an isothermal gas in equilibrium subject only to gravity, the last two terms in (29) vanish, and the first two terms on the right mutually cancel; the two constituents are distributed proportionately to e^{-h/H_1} and e^{-h/H_2}, corresponding to a concentration gradient with which is associated a downward velocity of diffusion of the lighter constituent, just balancing an upward velocity of diffusion due to the pressure gradient.

In general the conditions in the atmosphere are probably such that at most heights and times the velocity of diffusion does not vanish; certainly this is the case in the troposphere and lower stratosphere. The average composition of the air at any level is determined, however, by many other processes in addition to diffusion; besides those that specially affect the water-content, there is large-scale mixing by turbulent winds, and at some levels also there is dissociation or ionization by light. The net result depends on the relative intensity or speed at which these can modify the composition.

The speed of diffusion is governed by the various terms in the bracket on the right of (29), and by the value of D_{12}, the coefficient of diffusion, which is a factor affecting them all; D_{12} for any pair of constituents depends on their nature, and for a given pair it depends also on the temperature and density, approximately as T^s/ρ, where s does not differ greatly from unity. Some typical values of D_{12} (at 0° C. and atmospheric pressure) and s are as follows:

Gas pair	D_{12}			s
O_2—N_2	0·18	.	. .	0·79
CO_2—N_2	0·14			
CO_2—O_2	0·14			
H_2—N_2	0·67			
H_2—O_2	0·70	.	. .	0·76
He—A	0·64			
H_2—CO_2	0·55	.	. .	0·74

By far the most important variation of D_{12} in the atmosphere is due to its dependence on the density; this changes by a factor 10 in a

distance H' (§ 2), and therefore by many powers of 10 between the ground and, say, 100 km. height. Consequently, the expression for the rate of diffusion increases upwards, so far as D_{12} is concerned, at a considerable rate.

The observed constancy of composition in the troposphere indicates that there the mixing effect of turbulent winds altogether overcomes the separative tendency of diffusion. The ratio between the two effects is not known; we cannot yet estimate the rapidity of the mixing process even in the troposphere. Observations of visible atmospheric phenomena at various high levels, however, show that strong and variable winds occur there, as in the troposphere. We have therefore no valid reason to suppose that at other heights, where such visual evidence of winds may be absent, there is actually no wind and no mixing by wind.

The visual evidence of winds at high levels is of various kinds. Occasionally clouds (of unknown composition and origin) are observed, sometimes 'mother of pearl' at about 25 km. height (by day or night), and at other times 'luminous night clouds' at about 80 km. height (at night, when illumined by the beams of the sun while still below the ground horizon). Clouds of both these kinds [49–53] have been shown by Jesse and Störmer to have velocities of about 100 m./sec., predominantly from east to west over Europe. Again, long-enduring luminous trails of meteors or fire balls [58, 63] are occasionally seen to persist for some minutes (or even hours), and these usually become twisted in a way which seems to imply the existence of powerful winds, up to 200 km. per hour, at heights of 100 km. or so; these winds appear in some cases to vary considerably in direction or strength within a few kilometres range of height. The mixing power of such winds is at present, however, quite unknown.

In default of such knowledge it is not possible to calculate the composition of the air in the stratosphere, though the upward increase of D_{12} suggests that at some considerable height the gases may be able in some degree to separate out, the proportion of the lighter ones increasing upwards. There is indeed some observational evidence [6], obtained by analysis of air samples taken *in situ*, that this separation is already beginning at a height of about 20 km.; the helium volume-percentage, which retains its normal value $5 \cdot 27 \times 10^{-4}$ up to nearly 20 km., appears to be increased in the ratio $1 \cdot 05$ at about 22 or 23 km., and the oxygen percentage at the same level seems to be reduced below its tropospheric value, in the ratio $0 \cdot 99$. The change from the normal percentage is naturally greater for helium than for oxygen, since the

former departs far more than oxygen from the mean molecular weight of the air (and of course in the opposite sense).

Up to the greatest height (24 km.) at which the helium percentage has yet been measured, it remains extremely small, but the scale-height (§ 2) of helium so much exceeds that of oxygen and nitrogen that helium should become a main constituent at high levels, if diffusion had unhindered play from about 25 km. upwards—unless the still lighter gas, hydrogen, is also present. As regards hydrogen, its percentage amount in the troposphere seems to be one-tenth that of helium; in the region where an upward increase in the helium percentage has been observed, the proportion of hydrogen should increase upwards, at a still more rapid rate.

One opposing consideration, however, specially affects these two light gases in the outermost levels of the atmosphere; a proportion of the molecules of any atmospheric constituent will, by chance collisions, acquire velocities which exceed the value of 11 km./sec. necessary for escape from the gravitational field of the earth. Hence, in the outer layers where a molecule moving upwards would have little chance of colliding with other air molecules, such fast molecules will be lost altogether to the atmosphere [7]. This rate of escape of atmospheric constituents is greatest for the lightest gases, and depends considerably on the temperature. It seems possible that the outer layers are sufficiently hot, for at least an appreciable fraction of the time (e.g. on summer days) to make the rates of escape of helium and hydrogen of geophysical importance: so that any hydrogen and helium originally present may have leaked away in the course of the earth's long history, while the continual output of helium from the earth is balanced by an equal rate of escape from the top of the atmosphere.

15.4. Spectroscopic indications of the constitution. At present these considerations cannot be put to a quantitative test, but there is observational evidence which suggests that at no level in the atmosphere above 100 km. are hydrogen and helium predominant constituents. This indication is afforded by certain atmospheric spectra. Such spectra, both of emission and absorption, give valuable information bearing on the composition of the atmosphere.

The spectrum of the aurora, which is sometimes visible from about 100 km. up to a height of several hundred kilometres, establishes the presence, throughout that range of atmosphere, of molecular nitrogen and atomic oxygen. The spectrum of the night sky [18–20], caused by emission in the atmosphere at high but unknown levels, also indicates

the presence of atomic oxygen as a regular constituent at the levels concerned. Neither in these emission spectra, nor in the atmospheric absorption spectrum, is there any reliable indication of hydrogen and helium as important constituents at any height. The simplest and most natural (though not absolutely necessary) conclusion is that hydrogen is everywhere practically absent, and that the observed upward increase in the percentage helium content never raises this content to more than a few per cent. (at most), before the percentage begins to decrease upwards owing to the suggested escape of helium at the top of the atmosphere.

The absorption of sunlight in certain well-known bands of the solar spectrum, especially in the ultra-violet, is found to be due to ozone in the atmosphere. Through the work of Fabry, Dobson, Götz, Regener, and others [11–16], much is now known about the distribution and variations of the ozone content. The ozone percentage is everywhere very small; its maximum is at about 35 km. The actual ozone density has its maximum at about 22 km. The total amount at all heights, if collected at the bottom of the atmosphere under the pressure existing there, would form a layer 2 or 3 mm. thick. It undergoes interesting variations with the weather, and with the seasons (it has a maximum near the spring, and a minimum near the autumnal equinox). Its annual mean amount increases with the latitude. It seems to be both formed and destroyed by sunlight, its amount being determined by a slowly changing balance between several opposing processes. Chapman [15] constructed a theory of this equilibrium, from which he inferred that the ozone concentration must decrease rapidly upwards above the level of maximum, and that the process of dissociation of molecular oxygen, involved in the production of ozone, must be increasingly effective at great heights, where almost all the oxygen must exist in the atomic state. The theory does not indicate the lowest level at which this dissociation becomes practically complete; it may perhaps be about 100 km. [15]; Eropkin [16] has suggested a much lower level, 40 or 50 km. The qualitative conclusion as to the presence of atomic oxygen is in harmony with the evidence afforded by the spectra of the aurora and the night sky.

This discussion of the composition of the atmosphere may be briefly summed up by saying that probably nitrogen and oxygen are everywhere the two predominant constituents of the atmosphere. The proportion of molecular oxygen may decrease slightly in the stratosphere from about 20 km. upwards, but with increasing height the oxygen will be increasingly dissociated, and the proportion of atomic oxygen will

tend to increase upwards relatively to the molecular nitrogen; the latter is not readily dissociated by ultra-violet light, and there is little or no evidence for the presence of appreciable quantities of atomic nitrogen in the upper atmosphere except during aurorae.

15.5. The E and F layers of the ionosphere [21–34]. Sunlight of wave-length less than about 2,900 A (Ångström units, $1 \text{A} = 10^{-8}$ cm.) cannot be observed at the ground, owing to strong absorption in this spectral region by atmospheric ozone. A band of light of still shorter wave-length is absorbed by molecular oxygen, which is thereby dissociated, forming atomic oxygen and also ozone. The night-sky light is probably due to the slow conversion, into light, of the energy of dissociation of the oxygen, this energy being derived from the light absorbed by molecular oxygen during the day [14.72].

Still further in the ultra-violet, sunlight is capable of *ionizing* oxygen and nitrogen. No such light reaches the ground, but its emission from the sun, and absorption by the atmosphere, are indicated by the presence of several ionized layers there. Nitrogen is more readily ionized than dissociated by sunlight, so that the ions at high levels are probably those of molecular nitrogen, and molecular and atomic oxygen. These ions are likely to be relatively few in number compared with the corresponding neutral particles. The auroral spectrum is due mainly to emission by neutral and ionized molecular nitrogen, and by neutral atomic oxygen. Ionized atomic oxygen and nitrogen also appear to contribute to the auroral spectrum.

Radio methods of observation of the ionosphere have proved most successful in extending our knowledge of the upper atmosphere. They reveal two main layers, the lower (E) at about 120 km. height, and the other (F) at 200–350 km. The F layer is much thicker than the E layer, and consists of two parts, the lower called F_1 and the upper F_2; these are separate during the daytime in low latitudes and in summer; in higher latitudes and in winter and at night they are partly merged, so that the less intense lower layer F_1 becomes indistinguishable from, or appears only as an intensification on the underside of, the F_2 layer. Sometimes there is a 'sporadic' or 'intermediate' thin layer (E_2) in or just above the E layer. Below the E layer there is another (D) layer at about 70 km. height, which partly absorbs radio waves, and also reflects long waves. According to Maeda [29] a (G) layer, above, and more ionized than, the F_2 layer, can be observed at any time in the spring, autumn, and winter, but only by night during the summer.

The ionization in the E and F_1 layers has a daily maximum near noon,

and decreases throughout the night. Exceptional temporary increases of ionization (possibly sometimes due to thunderstorms or meteors) occur in the E layer, even at night. The ionization in the E and F_1 layers increases from winter to summer. It also varies in unison with the sunspot cycle, by 50 or 60 per cent., implying an increase of 130 to 160 per cent. in the ionizing radiation; the D, F_1, and F_2 layers also seem to undergo 11-year changes of ionization.

The daily variation of the F_2 layer shows a maximum near noon in winter, but in summer it has morning and afternoon maxima, and a minimum near noon. The seasonal variation shows maxima near the equinoxes and minima in summer and winter. There are much more pronounced day-to-day variations of ionization in this layer than in the E and F_1 layers. Attempts have been made to correlate such day-to-day variations with sunspottedness and with the magnetic activity. No clear day-to-day relations with sunspots have yet been established, but during magnetic storms [30] the F_2 layer (at night the F layer) seems to expand and move upwards, the electron density being reduced sometimes to one-quarter the normal amount. The F_1 layer is also somewhat reduced in intensity, without change of height. The E layer seems to remain unaffected.

Recently, strong evidence has accumulated showing that the very bright chromospheric eruptions occasionally observed on the sun (especially with the spectrohelioscope, using $H\alpha$ light), and lasting from a few minutes to an hour (10.3), are associated not only with a fleeting intensification of the S_q field over the sunlit hemisphere, but also with temporary abnormal ionization in the D layer below or at the base of the E layer. This extra ionization manifests itself by the fading of radio signals during the brief period of the eruption. The fading is due to *absorption* of the waves, which occurs by conversion of the energy of regular oscillation of the electrons, imposed on them by the electric waves, into thermal energy, at collisions with other particles. The absorption per particle is greater the more frequent the collisions. It is great in the region of short electronic free paths, when there are a sufficient number of electrons there. Usually waves of medium and short wave-length can pass upward through the D layer without serious absorption; but at the times of the radio 'fade-outs' the number of free electrons below the E layer is temporarily greatly increased [24], and absorption of the radio waves is complete before the waves penetrate this layer, or at least before they penetrate it twice (once travelling upward and once downward).

The solar outbursts which cause such fading may be situated almost anywhere on the visible hemisphere of the sun, even quite near the edge. The radio fade-out and the accompanying magnetic disturbance (a fleeting intensification of the S_q field) are confined to the sunlit hemisphere.

In high latitudes, in quiet magnetic conditions, the E and F layers are observed with the reduced intensity that would be expected from the low altitude of the sun in those regions, if their ionization is due to ultra-violet sunlight. But during and after times of great magnetic activity there is radio fading of a more local character than that described above; it is specially notable near the auroral zone [34]. The E and F layers then cease to be observable, owing to the formation of a strongly absorbing layer below them, due to the intense injection of electric particles which are supposed to come from the sun, and are guided by the earth's magnetic field to the higher latitudes (Ch. XXIV). They also produce the visible aurora, at heights down to 90 km. or less.

During solar eclipses [35–8] the ionization in the ionosphere decreases, in a way which shows that the main ionizing agent travels from the sun with approximately the speed of light; and it may therefore be concluded that the agent is ultra-violet light, and not neutral particles travelling with a distinctly smaller speed. It is not yet certain whether such a corpuscular ionizing agent may be an important though secondary cause of the F_2 layer and its variations (see Chapter XXIV).

Radio exploration of the ionosphere [2] yields (1) travel-times t' for reflected signals, (2) critical penetration frequencies f_P beyond which reflection ceases for each particular layer, (3) reflection-coefficients R. The equivalent height of reflection, h', is $c/\tfrac{1}{2}t'$, where c denotes the velocity of light; h' and R are measured as functions of the wave-frequency, and from them it is possible to estimate the approximate value of the average number of collisions per unit time made by an electron with the gas molecules; this is called the *collision frequency* (§ 9).

From f_P we determine $n_e/m_e + n_i/m_i$, where n_e or n_i is the number per c.c., and m_e or m_i is the mass, of electrons (e) or ions (i).

The earth's magnetic field in the ionosphere (*total* intensity H) polarizes the reflected radio waves [23], in a way which indicates that n_e/m_e is more important than n_i/m_i, or that n_e much exceeds $(m_e/m_i)n_i$. Thus the radio data indicate the value of n_e, or rather the maximum value of n_e in each layer. Molecular or atomic ˙ion-densities n_i go

undetected because they do not exceed the electron densities in the large ratio m_i/m_e, which for oxygen atoms is 3×10^4, and for nitrogen molecules is nearly twice as great.

The wave-frequency f (in cycles/sec., often stated in megacycles/sec. $= 10^6$ cycles/sec.) is given in terms of the wave-length λ by

$$\lambda f = c = 3 \cdot 00 \times 10^{10} \text{ cm./sec.} = 3 \times 10^5 \text{ km./sec.} \tag{30}$$

Let f, f_o, f_e denote the frequencies of the incident and of the *ordinary* and *extraordinary* reflected waves. Then

$$f^2 = f_o^2 = f_e^2 - f_e f_H, \tag{30 a}$$

where f_H is the electron gyro-frequency of § 10 ($= eH/2\pi m_e = \omega_e/2\pi$). Hence

$$H = (2\pi m_e/e)(f_e - f_o^2/f_e). \tag{30 b}$$

By observing f_e and f_o, Appleton [2] thus determined, by radio methods, the (total) magnetic intensity H in the F layer; his value was 10 per cent. less than the ground value, in reasonable accord with the inverse-cube law of variation of H with distance from the earth's centre.

The critical penetration-frequency f_{P_o} for the ordinary wave, for any reflecting layer, gives the maximum value of n_e for that layer,

$$n_e = 1 \cdot 24 \times 10^{-8} f_{P_o}^2. \tag{30 c}$$

The maximum electron-densities in the E and F_2 layers at the equator are thus found to be respectively about 2×10^5 and 10^6 per c.c.

From the measured dependence of the reflection-height h' on the frequency f, Appleton has deduced values of the scale-height H_0 (§ 2); in the E layer $H_0 = 11 \cdot 4$ km., over England; in the winter F layer H_0 is about 45 km., and in the F_2 layer in summer it is about 70 km. Budden, Ratcliffe, and Wilkes [22 i] have found the value $H_0 = 6$ km. at about 70 km. height, below the E layer. In the lower stratosphere, where $T = 220°$ absolute, $H_0 = 6 \cdot 4$ km. The higher values in the ionosphere must be due mainly to higher temperatures, though partly also to a reduction in the mean molecular weight.

From the observed dependence of the reflection-coefficient R on the frequency, Eckersley has determined the electron collision frequency ν_e in the F_1 layer, at an estimated height of 265 km., as $3 \cdot 6 \times 10^3$ sec.$^{-1}$ Farmer and Ratcliffe have similarly found $\nu_e = 1 \cdot 6 \times 10^3$ sec.$^{-1}$ for the region F_2. Appleton states that in the E layer ν_e is of order 10^4. At the ground the value of ν_e is about 2×10^{11}. Since ν_e is proportional (§ 9) to $n\sqrt{T}$, where n is the number-density of air molecules, and since $n = 3 \times 10^{19}$ at the ground, these values of ν_e imply that in the E layer

n is about 2×10^{12}, and in the F layer, of order 10^{11}—agreeing roughly with the above estimates of H_0.

From (30 c) it is clear that the reflections from the F_1 and F_2 layers are observable only because the higher layers have higher electron-densities and therefore reflect waves of higher frequencies (or shorter wave-lengths) that pass through the E layer. The ionization between the layers is not directly observable.

The actual radio-exploration is made by recording reflections from signals of various frequencies. In the automatic multifrequency technique, originated at the U.S. Bureau of Standards [24 a] and further developed by Berkner [22–4], successive echoes are recorded for signals emitted automatically at small increments of frequency over the range from 0·516 to 16·0 megacycles/sec. (wave-lengths 580 m. to 19 m.). Each 'sweep' through this frequency range is completed in 15 minutes, four sweeps being made each hour. In this way, according to (30 c), the least virtual height of each ion-density in the range $3·1 \times 10^3$ to $3·1 \times 10^6$ equivalent electrons per c.c. is measured, and from the curves thus obtained on the photographic trace, the critical frequency, minimum virtual height, and other characteristics of each region are determined.

The most complete ionospheric information so far assembled during a 24-hour period [10.11] is given in Fig. 1 for a station near Washington, D.C. It happens to include a widespread radio fade-out, simultaneous with an intense solar eruption. No changes in the F_1 and F_2 layers during the fade-out are noticeable, and the slight increase in the virtual height and maximum ion-density of the E region, just after the reflections became observable again after the fade-out, indicates that the effect occurs primarily below the 100 km. level, and that the waves are absorbed because of intense ionization in a region of high collision frequency. The stability of the F_1 and F_2 regions shows that they absorb only negligible amounts of the ionizing solar radiation producing the fade-out.

Fig. 1 a, after McNish and Ludy [11.15], shows the quality of transatlantic radio reception on individual days, 1928 May 28 to 1930 December 31, expressed as percentage commerciality (= the percentage fraction of time which is favourable for receiving); this is compared with the magnetic activity as measured by the American character-figure (increasing downward). The diagram shows clearly that quiet magnetic times are favourable for radio reception.

At present the existence of the several ionized layers, their location,

Fig. 1. The variations of the minimum virtual heights (A, above) and critical penetration frequencies (B, below) for the E, F_1, and F_2 regions of the ionosphere at 15-minute intervals on 1937 July 31 when a radio fade-out occurred, lasting from about 11^h 17^m ($75°$ W. meridian time) to 13^h 4^m. The 'normal' variations, as represented by the mean of the corresponding data for the 6-day period July 26–31, are given for comparison. (After L. V. Berkner and H. W. Wells)

intensity, variations, and other properties, cannot be inferred and calculated purely from known data, partly because they involve the intensity of the solar ionizing radiation. This cannot be found by direct observation (since this radiation is all absorbed high up in the atmosphere), and it is not yet deducible from theoretical solar physics.
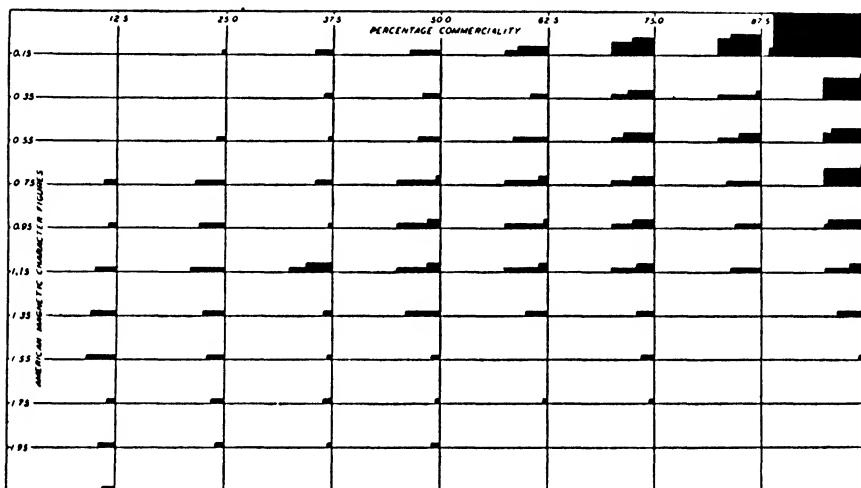

FIG. 1 *a*. Illustrating that magnetically calm days are usually days on which transatlantic radio reception is good, and that magnetically disturbed days are mostly also days of disturbed transmission. The magnetic activity is characterized by the American magnetic daily character-figure C_A, and the quality of the reception by the 'commerciality' P, being the percentage of the day during which the North American radio circuits were satisfactory for commercial use. The scales at the sides of the diagram indicate certain ranges of C_A and P, and in the corresponding rectangles of the Figure a space is blacked in, whose area is proportional to the number of days (within the period May 28, 1928 to December 31, 1930) characterized by values of C_A and P falling within these ranges. (After A. G. McNish and A. K. Ludy)

So far only a few tentative steps have been taken towards a theory of the ionosphere.

15.6. Ion-production by absorption of monochromatic light.
One such step has been to consider the simple case of ionization (or simple dissociation into neutral particles, as in the case of molecular oxygen) by absorption of monochromatic light (with a definite absorption coefficient A) in an atmosphere of uniform composition and temperature [39]. Such an atmosphere will be distributed exponentially, with a constant scale height H_0 (§ 2), here (§ 6) written H, without suffix.

Let S denote the intensity of such light or radiation, at height h above the earth, the intensity outside the atmosphere being S_∞. Let

ρ_0 be the atmospheric density at the ground, so that at height h, by (17), $\rho = \rho_0 e^{-h/H}$. Then when χ is the zenith-distance of the sun at the point and time considered, the decrease in S by absorption, during the path, of length $(\sec \chi)\, dh$, between the levels $h+dh$ and h, is given by

$$dS = -A\rho(\sec \chi . dh)S$$

or by

$$dS/S = -(A\rho_0 \sec \chi)e^{-h/H}\, dh,$$

which has the integral

$$\log_e S = -(A\rho_0 H \sec \chi)e^{-h/H}+\text{constant}.$$

Since $S \to S_\infty$ as $h \to \infty$, the constant of integration is $\log_e S_\infty$. Hence

[39] $$\log_e(S/S_\infty) = -(A\rho_0 H \sec \chi)e^{-h/H},$$

or $$S = S_\infty \exp(-A\rho_0 He^{-h/H} \sec \chi). \tag{31}$$

The absorption of the radiant energy per c.c. of atmosphere is $(dS/dh)\cos \chi$, and if β is the number of (simple) ions produced by the absorption of unit quantity of such energy, the rate of *production* of ions per c.c.—denoted by $I(\chi, h)$—will be $\beta(dS/dh)\cos \chi$, or, using (31),

$$I(\chi, h) = \beta A S_\infty \rho_0 \exp\{-h/H-A\rho_0 He^{-h/H} \sec \chi\}. \tag{32}$$

The total number of ions produced in a vertical column of air of 1 sq. cm. cross-section, by the complete absorption of the incident radiation, is $\beta S_\infty \cos \chi$.

The rate of ion-production $I(\chi, h)$ has a maximum value $I(\chi)$ at the height $h(\chi)$, where

$$\exp\frac{h(\chi)}{H} = A\rho_0 H \sec \chi, \tag{33}$$

$$I(\chi) = \beta S_\infty \cos \chi/H \exp 1. \tag{34}$$

Let the values of $h(\chi)$ and $I(\chi)$ for $\chi = 0$ (overhead sun) be denoted by h_0 and I_0. Then

$$e^{h_0/H} = A\rho_0 H, \qquad I_0 = \beta S_\infty/H \exp 1. \tag{35}$$

Consequently $$h(\chi) = h_0+H \log_e \sec \chi, \tag{36}$$

$$I(\chi) = I_0 \cos \chi. \tag{37}$$

In terms of I_0 and h_0, (32) may be written

$$I(\chi, h) = I_0 \exp\left\{\frac{h_0+H-h}{H} - \sec \chi \exp\frac{h_0-h}{H}\right\}. \tag{38}$$

These formulae can be expressed in a more concise and convenient form by measuring heights in terms of H as unit, from the height h_0 as datum. Let z denote the height so reckoned: it may conveniently

be called the *reduced* height, and may of course take negative values; clearly

$$z = \frac{h - h_0}{H}.$$ (39)

If $z(\chi)$ corresponds to $h(\chi)$,

$$z(\chi) = \log_e \sec \chi$$ (40)

and

$$I(\chi, h) = I_0 \exp(1 - z - e^{-z} \sec \chi).$$ (41)

In the above equations the curvature of the level surfaces in the
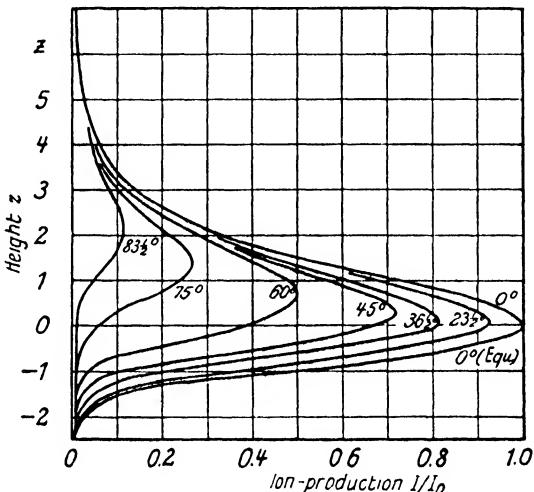
FIG. 2. The height-distribution of the rate of absorption of monochromatic radiation, or of the rate of ion-production, in a plane-stratified exponential atmosphere, for different zenith-distances of the sun. Heights (z) are indicated in terms of the scale-height as unit, relative to the level of maximum absorption when the sun is overhead

atmosphere has been neglected, that is, they have been treated as parallel planes. It has been shown that if the curvature is taken into account, the equations remain very nearly true provided that, wherever $\sec \chi$ occurs, it is replaced by a certain function $f(R, \chi)$. Here R denotes the distance from the centre of the earth to the height h_0 in the atmosphere, *expressed in terms of H as unit*. The function $f(R, \chi)$, when R is large, differs little from $\sec \chi$ until χ approaches and exceeds 90°; unlike $\sec \chi$, the function $f(R, \chi)$ remains finite for $\chi \geqslant 90°$. Chapman [39] has given a table of values of the function f. This modification of the above equations will not be further considered here.

Fig. 2 shows the distribution of ion-production with height, for

various zenith-distances χ; what is actually plotted is I/I_0, the curves being independent of the value of I_0. The decrease in the ion-production and the rise in its level of maximum, at low solar altitudes, is clearly shown. For overhead sun ($\chi = 0$), almost all the ion-production is confined within a range of z from $-1\cdot5$ to $3\cdot5$, or a height-range of $5H$. The ion-production falls away more rapidly below than above the level of maximum.

The sun's zenith-distance χ depends on the season, and on the latitude and time of day at the point considered. It is convenient to use the co-latitude θ instead of the latitude; θ varies from $0°$ at the north pole, through $90°$ at the equator, to $180°$ at the south pole. The season affects χ because it depends on the sun's north declination, which varies from $23\cdot5°$ at the northern summer solstice, through 0 at the equinoxes, to $-23\cdot5°$ at the northern winter solstice. The time of day, reckoned from local noon, may conveniently be expressed in seconds, when it will be denoted by t, or alternatively in angular measure (radians, or degrees) when it will be denoted by ϕ. When ϕ is expressed in radians, at the rate 2π per day,

$$t = 86400\phi/2\pi = 1\cdot37 \times 10^4\phi, \tag{42}$$

there being 86,400 seconds in a day.

Since the meridian through the sun corresponds to local noon ($\phi = 0$), the sun's zenith distance χ is given by

$$\cos\chi = \sin\delta\cos\theta + \cos\delta\sin\theta\cos\phi. \tag{43}$$

At the equinoxes $\delta = 0$ and

$$\cos\chi = \sin\theta\cos\phi \quad \text{(equinoxes)}. \tag{44}$$

At noon $\phi = 0$ and then

$$\chi = \tfrac{1}{2}\pi - (\theta+\delta) = \chi_0. \tag{45}$$

The variation of ion-production throughout the day, at any height, can be inferred from (41) and (43). It can conveniently be represented as a fraction f of the noon value, corresponding to $\chi = \chi_0$; the noon value itself is

$$I_0\exp(1-z-e^{-z}\sec\chi_0) = I_0\exp\{1-z-e^{-z}\operatorname{cosec}(\theta+\delta)\}; \tag{46}$$

its distribution with height, for various values of χ_0, is given by Fig. 2. The ratio f of the value at any time to that at noon, at the same season, place, and height, is

$$\exp[\{\operatorname{cosec}(\theta+\delta) - \sec\chi\}e^{-z}]. \tag{47}$$

At the equinoxes ($\delta = 0$) the ratio (47) may be expressed specially simply, as

$$\exp\{-\gamma(\sec\phi-1)\}, \tag{48}$$

where

$$\gamma = e^{-z}\operatorname{cosec}\theta. \tag{49}$$

At that season ϕ varies during the day hours from $-90°$ at sunrise to $90°$ at sunset. The proportionate daily variation of ion-production (taking that at noon as the unit) is illustrated by Fig. 3, for several values of γ. The following table gives the values of γ for various values of z and θ.

TABLE 3

$$\gamma = \operatorname{cosec}\theta \,.\, \exp(-z)$$

	θ							
z	$6\cdot5°$	$15°$	$30°$	$45°$	$53\cdot5°$	$60°$	$66\cdot5°$	$90°$ (eqr.)
4·0	0·162	0·071	0·037	0·026	0·023	0·021	0·020	0·018
3·0	0·440	0·192	0·100	0·070	0·062	0·057	0·054	0·050
2·0	1·196	0·523	0·271	0·191	0·168	0·156	0·148	0·135
1·5	1·971	0·862	0·446	0·316	0·278	0·258	0·243	0·223
1·0	3·250	1·421	0·736	0·520	0·458	0·425	0·401	0·368
0·5	5·358	2·343	1·213	0·858	0·755	0·700	0·661	0·607
0	8·834	3·864	2·000	1·414	1·244	1·155	1·090	1·000
−0·5	14·56	6·37	3·297	2·332	2·051	1·904	1·798	1·649
−0·75	18·70	8·18	4·23	2·99	2·63	2·445	2·308	2·117
−1·0	24·01	10·50	5·44	3·84	3·38	3·14	2·96	2·72
−1·25	30·83	13·49	6·98	4·94	4·34	4·03	3·81	3·49
−1·5	39·59	17·32	8·96	6·34	5·58	5·18	4·89	4·48

At the level of maximum ionization (h_0, or $z = 0$), at the equator ($\theta = \tfrac{1}{2}\pi$), $\gamma = 1$; the curve $\gamma = 1$ in Fig. 3 resembles, though except at $\phi = 0$ it is below, the curve of $\cos\phi$ (shown by a dotted line for comparison). For $\gamma > 1$, corresponding, *inter alia*, to levels below h_0 at the equator, or to levels somewhat above h_0 at higher latitudes, I/I_{noon} is still smaller, indicating that the part of the day during which the radiation is chiefly absorbed at those levels is more and more concentrated round the hour of noon; conversely for $\gamma < 1$, corresponding to the higher levels, the rate of absorption and ion-production remains more nearly constant throughout the day, the rise and fall near dawn and sunset being more rapid.

At seasons other than the equinoxes, when the sun's declination is δ, the proportionate daily variation of ion-production *at the equator* ($\theta = 90°$) can still be represented by a formula of the type (48), if γ now denotes $e^{-z}\sec\delta$; this means that the equatorial daily variation is then the same as the equinoctial daily variation at a station in latitude δ.

The slightness of the seasonal change in the proportionate daily variation of ion-production at any height above the equator is indicated, in conjunction with Fig. 3, by the small difference between corresponding values of γ in the last two columns of Table 3, which refer respectively, for any height z, at the equator, to the solstices and equinoxes.
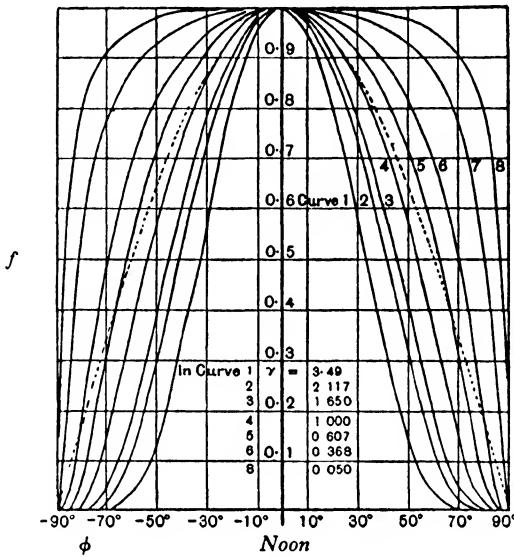

FIG. 3. The daily variation in the rate of absorption of monochromatic radiation, or in the rate of ion-production, in a plane-stratified exponential atmosphere, at the equinoxes. The different curves refer to different values of a parameter (γ) depending upon the co-latitude (θ) and the height (z); $\gamma = e^{-z}\cosec\theta$. The daily variation for each value of γ is indicated by the fraction (f) of the noon-value for that value of γ; the noon-value itself depends on the noon zenith-distance of the sun, which at the equinoxes is $90 - \theta^\circ$ (see Fig. 2). The time-scale refers to the sun's hour-angle ϕ reckoned forwards or backwards from noon, at the rate of 15° per hour

At latitudes far from the equator the proportionate daily variation of ion-production at any height changes considerably with the season, especially on account of the changing duration of the sunlight. This is illustrated in Figs. 4, 5, and 6, which refer to latitude 60° ($\theta = 30^\circ$) at the equinoxes, the summer solstice, and the winter solstice, for a series of heights between $z = -2$ and $z = 7$. The curve representing $\cos\chi$ is also shown by a dotted line on each diagram.

The different base-lengths in the three diagrams correspond, of course,
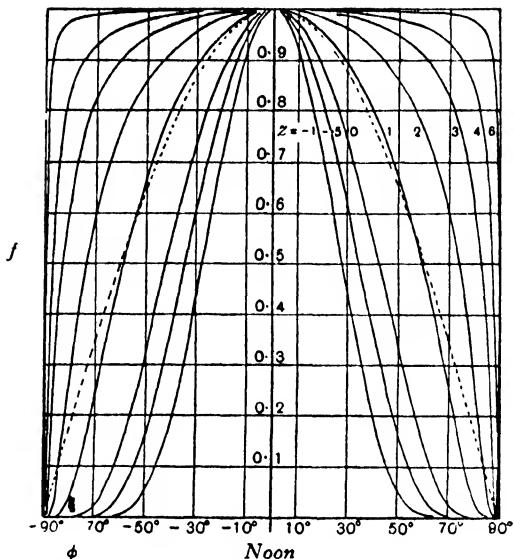
FIG. 4. The daily variation in the rate of absorption of monochromatic radiation, or in the rate of ion-production, in a plane-stratified exponential atmosphere, in latitude 60° at the equinoxes, at different heights z. The daily variation at each height is indicated by the fraction (f) of the noon-value at that height; the noon-value itself is different for different heights (see Fig. 2). The time-scale refers to the sun's hour-angle φ reckoned forward or backwards from noon, at the rate of 15° per hour
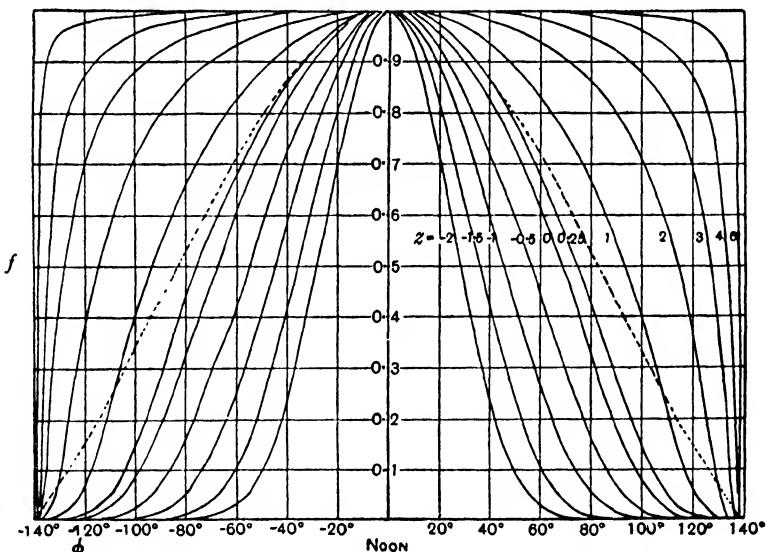

FIG. 5. Daily variation curves for latitude 60° (as in Fig. 4), at the summer solstice

to the different periods of daylight at the three seasons. In summer, when the sun's rays are most direct, the rate of ion-production at the level $z = 0$ exceeds half its maximum (noon) value for about half the period of daylight; this is nearly true at the same level at the equinoxes also, but in winter the absorption of radiation at this level increases more slowly, and it exceeds half the noon value only for about a quarter of the period of daylight. At $z = 6$ at all seasons the absorption is nearly at its full value throughout almost the whole day.

To obtain a comprehensive idea of the variation of ion-production with respect to season, height, latitude, and local time the curves in Fig. 2 should be considered in conjunction with those of Figs. 3–6.

This discussion of light-absorption and ion-production is simplified by assuming that the atmosphere (or at least the absorbing constituents in it) is distributed exponentially with height, and that the ionizing radiation is purely monochromatic. In the earth's ionosphere neither condition is likely to be strictly fulfilled. The scale-height H will not be strictly constant, and on this account the height-distribution curves will not exactly resemble those of Fig. 2, though they are likely to retain the same general form. The ionizing radiation will also be distributed over


FIG. 6. Daily variation curves for latitude 60° (as in Fig. 4), at the winter solstice

a certain band of wave-length, with a corresponding range in the coefficient of absorption A. Each monochromatic constituent would, in a truly exponential atmosphere, be associated with a layer of absorption and ion-production such as is illustrated in Fig. 2, but the height and the value of the maximum ion-production would be different for different values of A. The resulting curve of height-distribution of ion-production would be obtainable by summing the ion-production in all the overlapping layers. It would still somewhat resemble the curves in Fig. 2, but would be more spread out, mainly below the level of maximum. This has been confirmed by Chapman [75], who considered a particular form of absorption band.

The existence of more than one ionized layer, and of the ozone layer, shows that different components of the sun's ultra-violet radiation have

widely different absorption coefficients. The height-curves of ion-production (or oxygen-atom production) may be expected to have similar forms in each case, though their height-scales will differ from layer to layer, because H has seriously different values in the E and F (and ozone) layers.

Radio measurements indicate that the E layer is thin and the F layer thick, since at 120 km. H is much smaller than at 250 or 300 km. (p. 501).

The small thickness of the E layer has recently been attributed, independently by Wulf and Deming [73], and by Mitra [74], to its being a layer of ionized molecular oxygen, situated at the transitional region where the degree of oxygen dissociation increases rapidly upwards. This explanation, though it naturally involves uncertain assumptions, seems promising. As regards the F layers, Wulf and Deming attribute both of them to ionization by radiation corresponding to two different absorption bands of molecular nitrogen, whereas Mitra attributes only the F_1 layer to nitrogen; he ascribes the F_2 layer to atomic oxygen.

Saha has suggested that monochromatic Lyman α radiation (from solar atomic hydrogen) may be responsible for the temporary intense ionization below the E layer, which causes radio fade-outs. Thunderstorms and meteors may produce fleeting E layer ionization [83, 84].

The contribution made by cosmic rays to the ionization of the E and F layers seems quite negligible compared with that due to sunlight [85].

15.7. The distribution and variations of ion-density. The ion-production I, considered in § 6, is to be carefully distinguished from the ion-density n_i; the former is the number *produced* per second per c.c., and the latter the number per c.c. *existing* at any time. The calculation of the former, under the reasonable simplifying assumptions of § 6, is mathematically easy. The calculation of n_i involves further physical hypotheses, some of which are at present doubtful, and the results cannot be expressed in analytically simple formulae; they are best shown by a series of graphs.

Let n_e, n_+, n_-, n_i denote respectively the number of electrons, and of positive, negative, or positive *and* negative ions, per c.c. Then

$$n_i = n_+ + n_-, \tag{50}$$

and, if the ions are all singly charged and the air is neutral,

$$n_e + n_- = n_+. \tag{51}$$

The negative ions are formed by attachment of electrons to neutral particles.

The changing balance between ion (and electron) production, electron attachment and detachment, and recombination of positive ions with electrons or negative ions, is represented by the equations

$$\frac{dn_e}{dt} = I + \delta n_a n_- - \alpha n_+ n_e - \alpha_e n_a n_e, \qquad (52)$$

$$\frac{dn_+}{dt} = I - \alpha n_+ n_e - \alpha_\pm n_+ n_-, \qquad (53)$$

$$\frac{dn_-}{dt} = \alpha_e n_a n_e - \alpha_\pm n_+ n_- - \delta n_a n_-, \qquad (54)$$

where I denotes the number of ions (or electrons) produced per c.c. per sec., that is, the ion-production (§ 6), n_a denotes the number of *neutral* particles per c.c., of the kind to which electrons can attach themselves, and δ, α_e, α, $\alpha_\pm$ are factors called detachment, attachment, or recombination coefficients. These equations indicate that the rate of change of n_e, n_+, n_- is due to processes of formation (at rates I, $\delta n_a n_-$, and $\alpha_e n_a n_e$) and dissipation by attachment, detachment, or recombination.

These equations will first be considered in connexion with the simplifying assumption that

$$\alpha_e = 0, \qquad (55)$$

that is, no electron attachment, so that $n_- = 0$, and consequently, by (51),

$$n_e = n_+. \qquad (56)$$

Hence it is convenient to use n to denote either n_+ or n_e (note that this usage differs from that of § 2). We have now only one equation to consider,

$$\frac{dn}{dt} = I - \alpha n^2. \qquad (57)$$

Here, as has been shown in § 6, I (which is supposed zero during the night) is, during the day, a function of the height h (or z), of the time of day t or ϕ, of the co-latitude θ, and of the season (through the sun's declination δ); α may also be variable with some of these factors, but will here be supposed constant, independent of the level z and of any of the variables θ, ϕ, δ.

By substituting for I from (41), and replacing t by ϕ according to (42), we may express (57) in the form

$$\sigma_0 \frac{dv}{d\phi} + v^2 = \exp(1 - z - e^{-z} \sec \chi) \quad \text{(day)} \qquad (58\,\text{a})$$

$$= 0 \quad \text{(night)}, \qquad (58\,\text{b})$$

where
$$\nu = n/n_0, \qquad (59)$$

$$n_0 = (I_0/\alpha)^{\frac{1}{2}}, \qquad (60)$$

$$1/\sigma_0 = 1.37 \times 10^4 (I_0 \alpha)^{\frac{1}{2}}. \qquad (61)$$

Clearly n_0 is the steady value which n would have at the level h_0 (or $z = 0$) at the equator at midday, if the earth did not revolve, i.e. if the absorption of radiation preserved its noon value there without any change with time; this may be seen from (57) by putting $dn/dt = 0$ (since the supposed state is steady) and $I = I_0$. Hence ν signifies the ratio of the ion-density n at any place and time, to the maximum possible ion-density which the given radiation could produce in the given atmosphere. The quantity σ_0 is a parameter entering into the transformed equation (58) giving ν. Evidently I_0 and α are expressible in terms of n_0 and σ_0:

$$1/\alpha = 1.37 \times 10^4 \sigma_0 n_0, \qquad (62)$$

$$I_0 = n_0/(1.37 \times 10^4 \sigma_0). \qquad (63)$$

The variation of ν during the night, expressed by (58 b), is easily seen to be

$$\nu = \frac{\sigma_0}{\phi + C} \quad \text{(night)}, \qquad (64)$$

where C is a constant of integration. The combined equations (58) have to be solved subject to the condition that ν is a purely periodic function; suppose ν_r and ν_s are its values at sunrise and sunset, and ϕ_r, ϕ_s the corresponding values of ϕ. Then from (64), expressed for these two epochs, C can be eliminated, giving the relation

$$\frac{1}{\nu_r} - \frac{1}{\nu_s} = \frac{\phi_r - \phi_s}{\sigma_0}. \qquad (65)$$

The solution of (58 a) has to be found subject to this condition. By graphical or numerical methods the equation can be solved without difficulty. Millington [40] has described a convenient method for the purpose.

Figs. 7–9 [39] give some graphs illustrating the daily variation of n/n_0 for different heights, at the equator, at the equinoxes, for three chosen values of σ_0, namely 1, $\frac{1}{5}$, $\frac{1}{25}$; Figs. 10–12 give corresponding curves for $\sigma_0 = \frac{1}{25}$, for the latitude 60°, at the equinox and the two solstices. (In Figs. 7–10, ϕ is reckoned from sunrise instead of from noon as in the other figures here given.) For $\sigma_0 = 1$ the maximum value of n/n_0 occurs distinctly after noon at heights adjacent to h_0, and at greater heights the maximum is deferred till near sunset. But for $\sigma_0 = \frac{1}{5}$
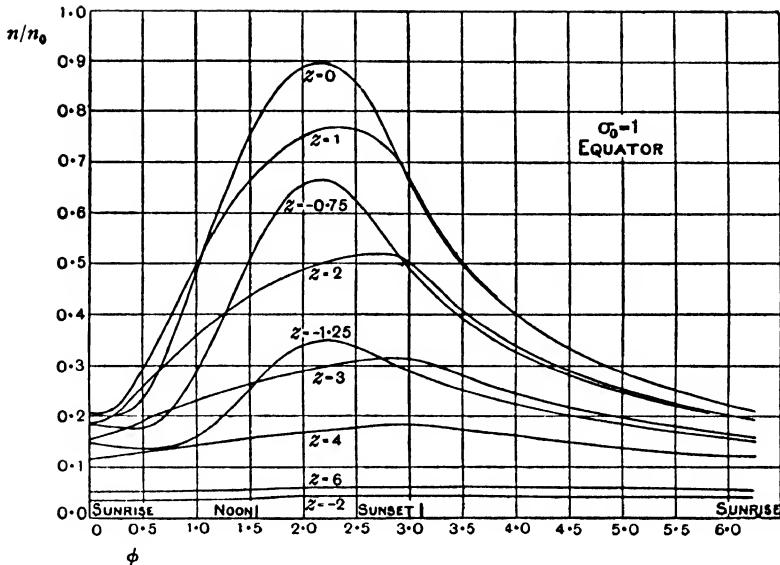
FIG. 7. Curves showing the daily variation of ion-density or electron-density at different heights (z), at the equator at the equinox, due to the absorption of monochromatic ionizing radiation in a plane-stratified exponential atmosphere, assuming simple recombination by two-body collisions. The densities (n) are expressed as fractions (n/n_0) of the maximum density (n_0) corresponding to a stationary overhead sun. The time-scale refers to the sun's hour-angle ϕ reckoned in radians from ground sunrise. The daily variations depend on a parameter σ_0, here taken as 1
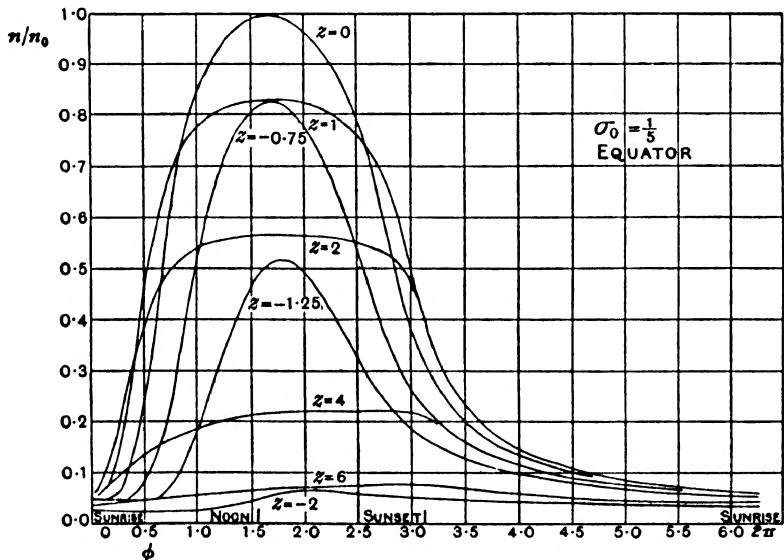

FIG. 8. Daily variation curves as in Fig. 7, for the equator at the equinox, for $\sigma_0 = \frac{1}{5}$
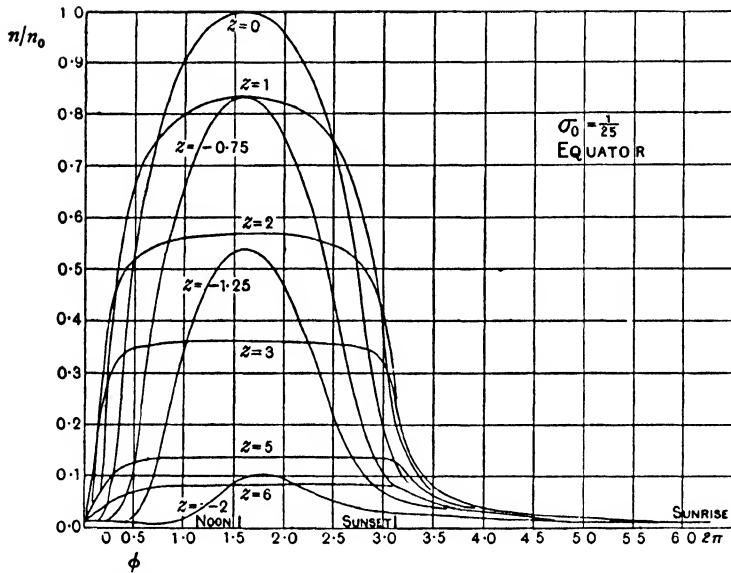
FIG. 9. Daily variation curves as in Fig. 7, for the equator at the equinox,
for $\sigma_0 = \frac{1}{25}$
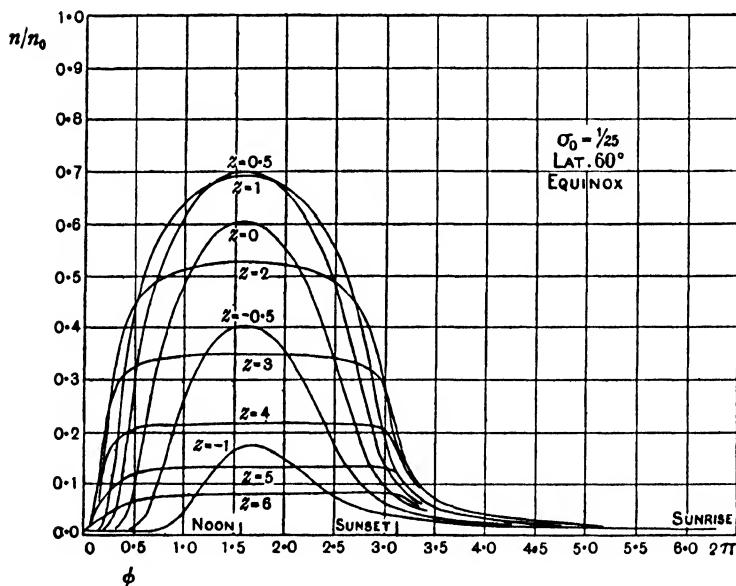
FIG. 10. Daily variation curves as in Fig. 7, for latitude 60° at the equinox,
for $\sigma_0 = \frac{1}{25}$
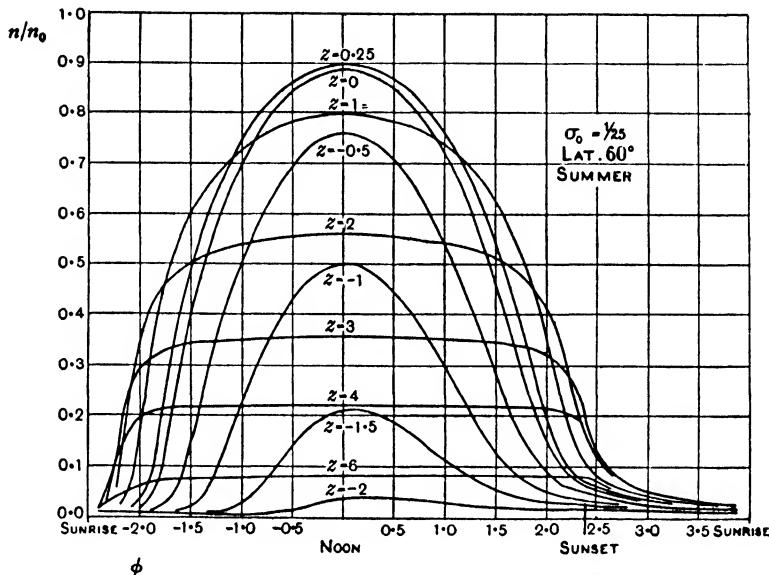
FIG. 11. Daily variation curves as in Fig. 7, for latitude 60° at the summer solstice, for $\sigma_0 = \frac{1}{25}$
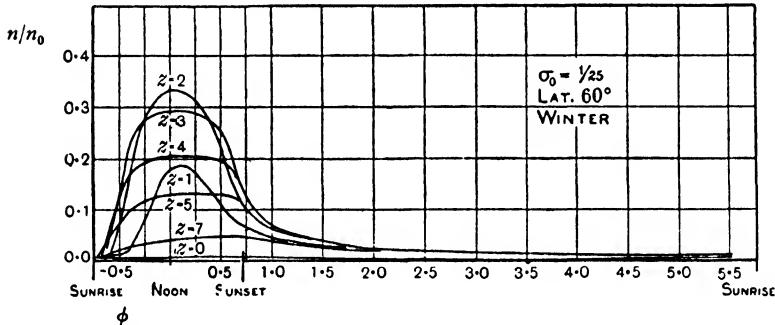
FIG. 12. Daily variation curves as in Fig. 7, for latitude 60° at the winter solstice, for $\sigma_0 = \frac{1}{25}$

and $\sigma_0 = \frac{1}{25}$ the noon value is very near the maximum value; this is true likewise for $\sigma_0 = \frac{1}{25}$ at latitude 60°, and therefore also at intermediate latitudes; in such cases the noon value of n/n_0 must be very nearly equal to the noon value of $\sqrt{(I/I_0)}$.

The seasonal variation of the daily maximum of n/n_0 at any level is very small at the equator; at 60° latitude, it is considerable at the lower levels, but much less at $z = 6$ and above.

The results of Figs. 7–12 can also be represented in another way, by

PROFESSIONAL CONDUCT
AND ADVOCACY
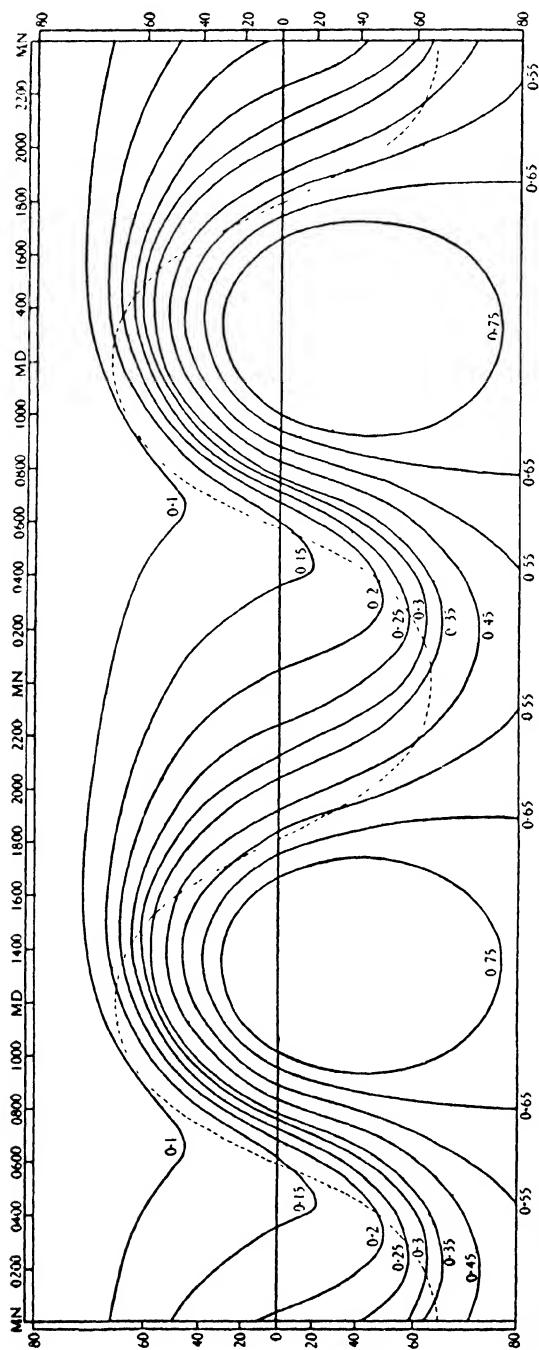
FIG. 14. The world-distribution of maximum ion-density or electron-density, as in Fig. 13, for $\sigma_0 = \frac{1}{2}$, for the northern winter solstice. The dotted line indicates the boundary (a great circle) between the sunlit and dark areas. (After G. Millington)
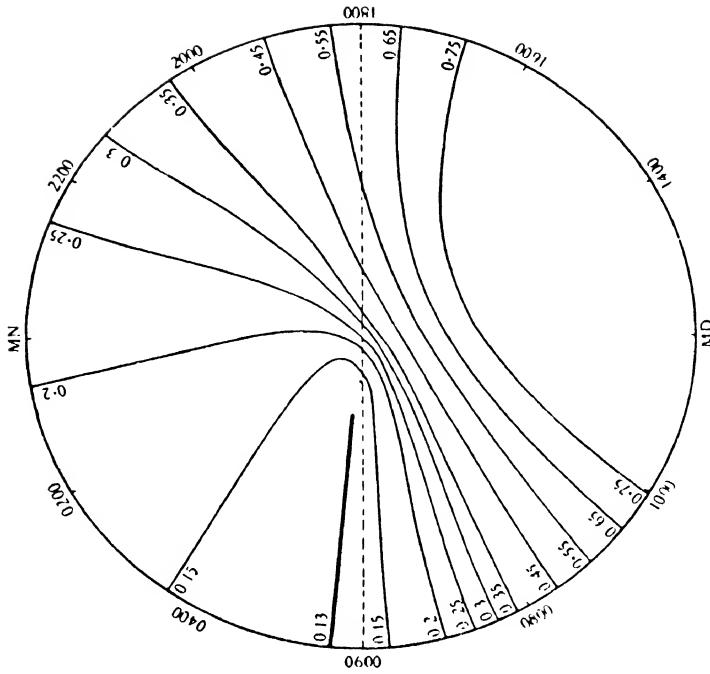
Fig. 16
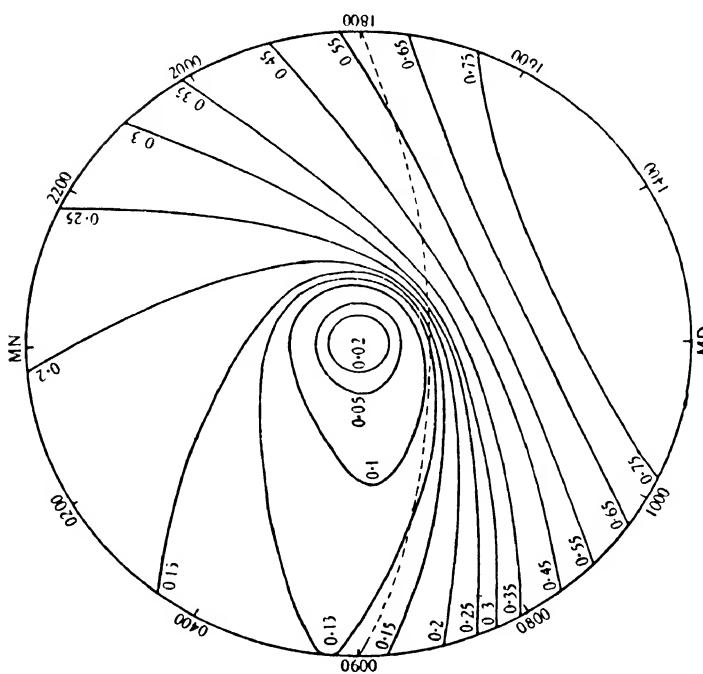
Fig. 15

Figs 15 and 16. The distribution of maximum ion-density or electron-density, as in Fig. 13, over the northern hemisphere, for $\sigma_0 = \frac{1}{2}$, at the winter solstice (Fig. 15), at the equinox (Fig. 16), and at the summer solstice (Fig. 17). (After G. Millington)
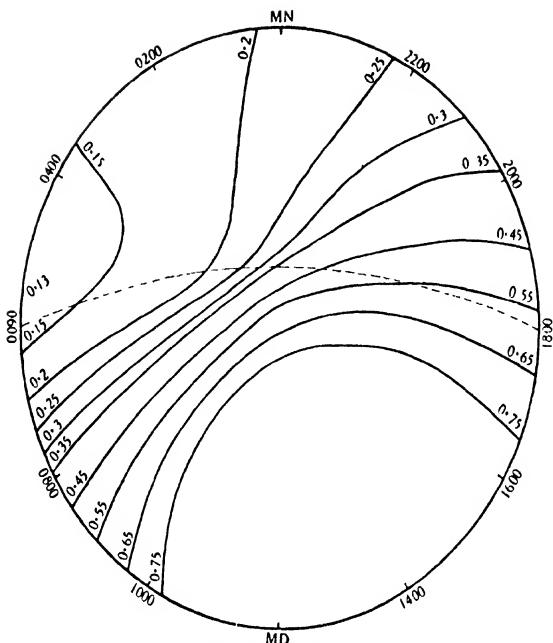
Fig 17 (see p. 521)
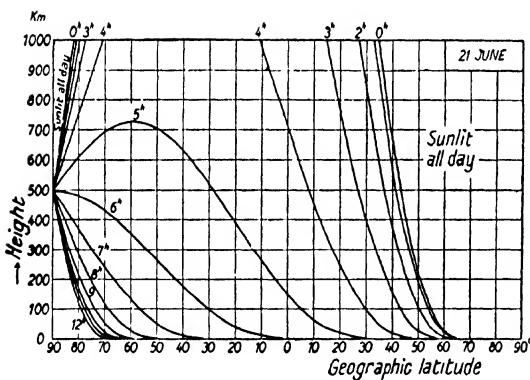
Fig. 18. Times of sunrise at different heights in the atmosphere in different latitudes at the summer solstice; the lines connect the points at which the sun rises at the times indicated
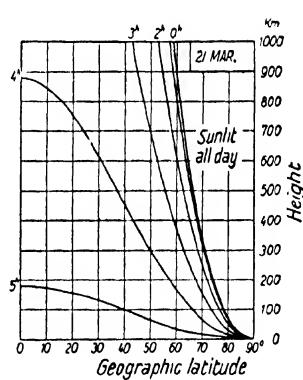
Fig. 19. Times of sunrise at different heights in the atmosphere in different latitudes, at the equinox; the lines connect the points at which the sun rises at the times indicated
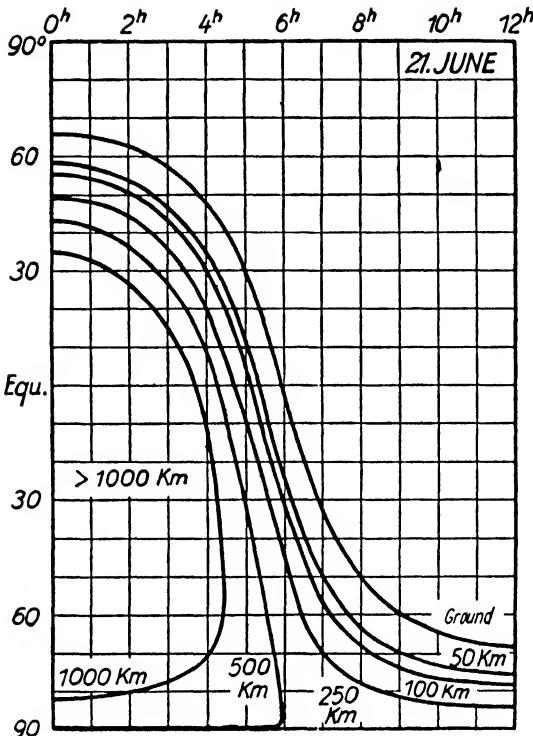
Fig. 20
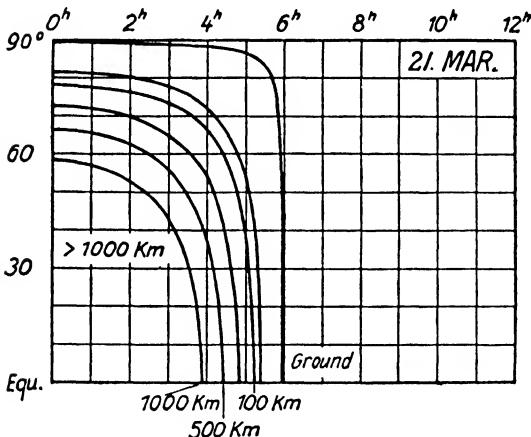
Fig 21

Figs. 20 and 21. Charts showing the places at which, at the ground and at different heights in the atmosphere, the sun is just rising, at the northern summer solstice (above) and the equinox (below). The meridians (between midday and midnight) are indicated by their true local times
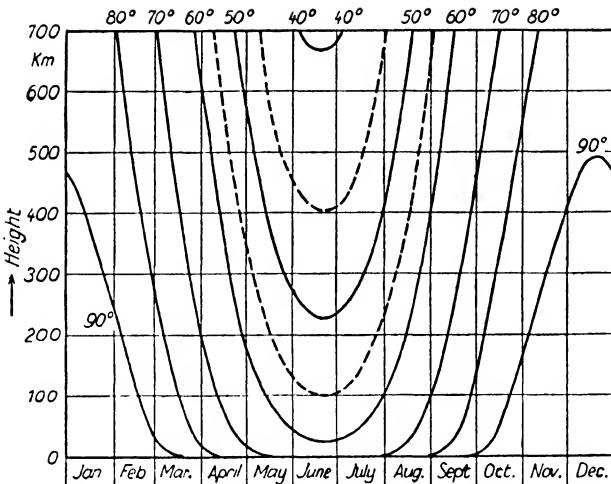
FIG. 22. Lines showing the heights above which, at different times of the year, and at various latitudes (40° to 90°), the atmosphere is sunlit at midnight
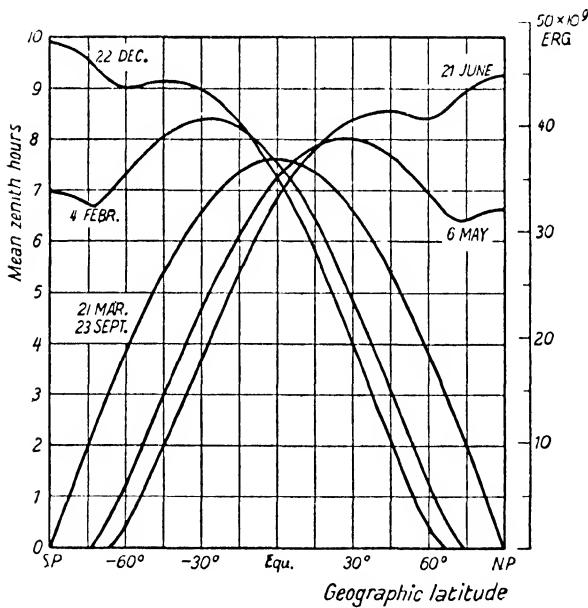
FIG. 23. The total solar radiation received in the course of a day, on a horizontal area of 1 sq. cm. just outside our atmosphere, in different latitudes and at different dates (taking account of the sun's changing distance). The amount is indicated on the right in ergs (taking the 'solar constant' to be 1·93 cal./cm.² min. or 1·35 × 10⁶ erg/cm.² sec.), and on the left by the number of hours required to give the same supply if the sun remained steady in the zenith at its mean distance

the sun's radiation is constant, the varying distance of the earth from the sun (5.1) will cause S_∞ to be approximately 6 per cent. greater at perihelion (January) than at aphelion (July). But account has been taken, in (66), of the possibility that α and T may vary throughout the year; the atmospheric temperature T is involved in n because it occurs in H, as in (9). If α and T are constant, (66) reduces to

$$\frac{(n_{max})_s}{(n_{max})_w} = \sqrt{\left(\frac{\sin(\theta + 23 \cdot 5°)}{\sin(\theta - 23 \cdot 5°)}\right)}. \tag{67}$$

Measurements for the E and F_1 regions, made at Slough (England), and at Deal, N.J. and Washington (in the U.S.A.), are in fair agreement with this simple relation.

In the F_2 region, on the contrary, the daily maximum value n_{max} appears to have a minimum value in summer instead of in winter, according to observations in America [31]. Suggested explanations of this remarkable result have been given by Hulburt [32] and Appleton [33], based on a supposed heating of the atmosphere by the absorbed solar radiation: this heating would expand the atmosphere and consequently reduce the electron density by increasing the volume within which the pre-existing electrons are distributed: it would also reduce the maximum ion-production $I(\chi)$, given by (37), through an increase in the factor H in I_0 (cf. (35)). If the latter were the only operative cause, the observed F_2 summer and winter results for England would require $T_s/T_w = 3 \cdot 4$. However, subsequent observations of n_{max} in the F_2 layer at Watheroo, in the S. hemisphere, indicate a different seasonal variation there [22 g].

It is not to be expected that the simple theory of §§ 6, 7 can account for all the features of the various ionized layers. The existence of separate or partly separate ionized layers E, F_1, and F_2 seems to imply the absorption of at least three spectral bands of ultra-violet radiation, having different absorption coefficients A; the absorbing constituents of the atmosphere in the layers may also be different, possibly partly dependent on the existence, during the day, in proportions varying with the level, of various excited states of molecular and atomic oxygen. The simple theory takes no account of changes in atmospheric temperature with height or time (which would modify H), nor of diffusion, nor of electron attachment. The constancy of the recombination coefficient α is also open to doubt; Pedersen [21] has given reasons for supposing that α is the sum of two parts, one proportional to the density of the atmosphere and the other independent of the density, the former

corresponding to recombination by three-body encounters, and the latter to simple recombination with emission of radiation. The two terms will probably be equal at some level in the ionosphere; somewhat above that level the constant term will become predominant, and there the simple theory might be valid except for variations of T and electron attachment and detachment.

In the E layer, where (57) seems to be valid for n_e, the temperature variations are perhaps unimportant; but (57) is satisfied there, not because $\alpha_e = 0$, but because the simple recombination conditions are simulated by more complicated processes. The electrons are removed from participation in the reflection of radio waves as effectively by attachment to neutral particles as by recombination with positive ions. Since neutral particles greatly outnumber the positive ions, the chance of attachment of an electron to a molecule, per impact, would have to be extremely small for the attachment rate to be negligible. But in addition to attachment, there is detachment of electrons by collision of neutral particles with negative ions; the importance of such detachments was suggested by Martyn and Pulley [71]. Massey [77] and his fellow-workers have determined the probability of the recombination, attachment, and detachment processes per collision, using the quantum theory. Their results suggest that, in the E layer, attachment and detachment proceed much faster than recombination, so that a dynamic equilibrium between electron attachment and detachment is maintained. From (54), neglecting the small terms dn_-/dt and $\alpha_+ n_+ n_-$, we infer that $n_-/n_e = \alpha_e/\delta \equiv \lambda$, so that, by (51), $n_+ = (1+\lambda)n_e$. Consequently (53) may be written

$$\frac{dn_e}{dt} = \frac{I}{1+\lambda} - (\alpha + \lambda \alpha_\pm)n_e^2 = I' - \alpha' n_e^2,$$

which is similar to (57), but with I' and α' substituted for I and α.

The values of α' deduced from the daily variation of n_e for the E layer are 12×10^{-9} by day and 4×10^{-9} by night; Bates, Buckingham, Massey, and Unwin [77 a] have shown that both α and $\alpha_\pm$ are much smaller than these values of α', and it is necessary to suppose that α' is approximately equal to $\lambda \alpha_\pm$, and that λ is of the order 100.

Best and Ratcliffe [79] and Appleton and Sayers [80] have shown that α' is constant over a range of pressure at least in the ratio 4 to 1. This is consistent with the view that the recombination of positive and negative ions is by a reaction such as

$$O^+ + O^- = O' + O'',$$
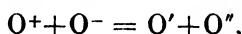
which is independent of the pressure; and Appleton and Sayers have shown that the part of α_+ associated with three-body collisions (which is proportional to p) must be small compared with the part associated with two-body collisions (which is independent of p); the former is $8 \cdot 5 \times 10^{-12}$ at a pressure of 10^{-3} mm.; the latter is approximately $2 \cdot 3 \times 10^{-8} P(\alpha)$, where $P(\alpha)$ is the probability of recombination per two-body collision between a positive and a negative ion. Hence $P(\alpha)$ can be as low as 10^{-3} without a_+ being seriously dependent on p in the E and F_1 layers.

In the F_2 layer the value of α inferred from the radio observations is of the order 10^{-10}, and there λ must be of the order 1 or less.

In the remaining part of this chapter the symbol H' will be used to denote the decimal scale-height, and the symbol H will denote the (whole) intensity of the magnetic field at any point in the atmosphere.

15.9. The atmospheric temperature, the mean free path, and the collision frequency. The radio methods of observation pick out certain features of the upper atmosphere, about which they give definite and valuable information; this stands out like a high light in a picture with a dark background. Other features of the picture, lying in this obscure background, must be known if the radio data are to acquire their full value for the understanding of the upper atmosphere and the geomagnetic phenomena therein produced. One factor on which much depends is the density of the atmosphere which, in turn, depends on the distribution of temperature and composition all the way up from the ground. The composition in the ionosphere, which has already been discussed in § 4, depends greatly on the height, as yet unknown, at which the oxygen first becomes nearly completely dissociated.

The temperature, which at about 20 km. is about 220° absolute (in temperate latitudes), appears to increase upwards from some rather greater height (perhaps 30 or 35 km.), to about 300 or 400° absolute at 40 or 50 km. height. Three lines of evidence suggest this: firstly, the theory of meteors by Lindemann and Dobson [54–8]; secondly, the abnormal propagation of sound to great distances, by downward refraction in the layer of upward increasing temperature [59–63]; and thirdly, the theory of the solar semi-diurnal variation of barometric pressure [64–9], as recently extended by Taylor [64] and Pekeris [65]. Eropkin, however, has suggested that the sound-propagation effects usually attributed to a rise in temperature may be largely due to an upward decrease of mean molecular weight m of the air, consequent on nearly complete dissociation of molecular oxygen being attained at a

level of 30 km. or so. (The velocity of sound is $\sqrt{(\gamma k T/m)}$, where γ is the ratio of the specific heats at constant pressure and volume respectively; a change in this velocity might be due to a change either in T or m.)

This relatively hot (or light) layer will cause the upward decrease of atmospheric density to be less rapid than it would be if the temperature and composition of the stratosphere did not vary with height. It is not known, however, how thick is the supposed hot layer. Vegard [14.60] concludes from a study of the bands in the auroral spectrum that at
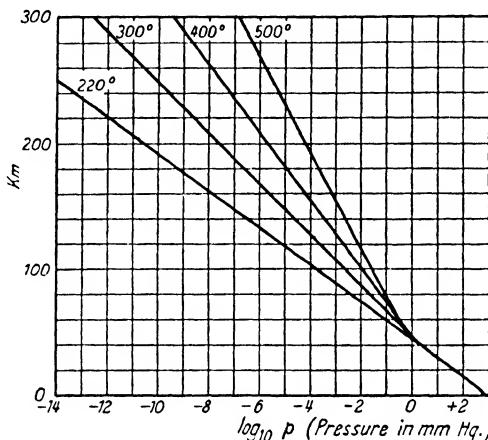

FIG. 24. The variation of the pressure with the height, in an atmosphere everywhere of the same composition as at the ground, for different assumed distributions of temperature

about 100 km. height the temperature is about 240° absolute, not much more than in the lower stratosphere. If this be so, Appleton's value (11·4 km.) of H_0 at 120 km. (§ 5) would imply a low value of m (that is, much dissociation). Pekeris [65] also found it necessary, in his work on atmospheric oscillations, to postulate an upward decrease of temperature above the hot layer, towards the 100-km. level.

At still higher levels it is possible that, at least in the daytime in summer, the temperature attains much higher values, perhaps approaching 1,000°. This idea, first proposed by Maris [9] and Hulburt [10], has lately found some support from the anomalous daily and seasonal changes of electron density in the F_2 layer, as interpreted by Hulburt [32] and Appleton [33].

The uncertainties as to the value of T at heights above 20 km. render it difficult to estimate the air density or pressure at great heights. This is illustrated [3] by Fig. 24, which shows the height-distribution of

$\log_{10} p$, where p is the pressure (there expressed in mm. of mercury), when the air composition is supposed to be everywhere the same as in the troposphere, and T in the stratosphere is supposed to have the value 220° at all heights or, above 45 km., one of the values 300°, 400°, or 500°. The dissociation of oxygen, by lowering the mean molecular weight, and still more the presence of helium or hydrogen in important proportions, would increase the slope of the lines in this figure, in the region concerned; this would correspond to still greater pressures at the same height. If in the outermost layer the gas is almost completely ionized, its upward decrease of density with height is further reduced; it would correspond to that for a gas with a molecular or atomic weight equal to half that of the neutral gas.

The uncertainty of the theoretical calculations of the pressure and density of the atmosphere at great heights also affects the calculation of the mean free path l, and of the collision interval τ for molecules or electrons, at great heights. They are connected with the number-density n (treating the air as a simple gas composed of spherical molecules of diameter d) by the relations

$$l = \frac{1}{\sqrt{2}} \frac{1}{n\pi d^2}, \tag{68}$$

$$\tau = \frac{l}{v} = \frac{1}{\sqrt{2}} \frac{1}{n\pi \, d^2 v} = \frac{1}{4n \, d^2} \sqrt{\frac{m}{\pi kT}} = \frac{1}{\nu}, \tag{69}$$

where v is the mean molecular speed, itself dependent upon the temperature T and the molecular mass m:

$$v = \sqrt{\frac{8kT}{\pi m}}. \tag{70}$$

The collision frequency of a molecule, that is, the number of collisions it makes per second, is $1/\tau$; it is usually denoted by ν, and this symbol will here be used in this sense, not in the sense of § 8.

Measurements of the reflection coefficient R of the ionosphere for radio waves afford information as to the electron collision frequency ν_e: in fact

$$-\frac{\partial}{\partial t} \log R = \frac{\nu_e}{2c} \frac{\partial}{\partial t} (P' - P),$$

where t denotes the time, c denotes the velocity of light, and P, P' denote the 'equivalent' and 'optical' paths of the waves. By measuring corresponding time-changes of $\log R$ and P', and neglecting changes of P, Eckersley has found that in the F_1 layer, at an estimated height of 265 km., $\nu_e = 3 \cdot 6 \times 10^3$, $\tau_e = 2 \cdot 8 \times 10^{-4}$ sec.

Farmer and Ratcliffe, using a similar method, have found the value $\nu_e = 1{\cdot}6 \times 10^3$, $\tau_e = 6{\cdot}2 \times 10^{-4}$ sec. for the higher region F_2. From the latter, by means of (69), taking $d = 10^{-8}$ cm., $T = 500°$ K.,

$$m_e = 9{\cdot}0 \times 10^{-28} \text{ gm.}, \qquad v_e = 1{\cdot}4 \times 10^7 \text{ cm./sec.},$$

the number-density of the air at the relevant height is inferred to be $2{\cdot}6 \times 10^{11}$.

This estimate of n will not be much affected by any likely change of T at the height concerned; the assumed value of d is more important, since if we took $d = 10^{-7}$ cm., the estimate of n would be reduced a hundredfold, to $2{\cdot}6 \times 10^9$. In applying (69) to electrons in a gas, d is usually interpreted as the mean of the diameters of the electrons and gas molecules, and the electronic diameter is taken to be negligible, so that the d used is half that for a gas molecule; $d = 10^{-8}$ is a reasonable estimate on this basis, but it is possible that the electron diameter (or the extension of its field of force) should not be neglected, in which case a rather larger value of d (though probably not more than 3×10^{-8}) should be used. In any case, if the 'experimental' value of τ here considered is fairly accurate, n is greatly in excess, even in the F_2 layer, of the electronic number density. A value of n of the order 10^{11}, for a height of, say, 300 km., when compared with the known value of n at 20 km. height, which is $1{\cdot}9 \times 10^{18}$ or $10^{18{\cdot}3}$, corresponds to $18{\cdot}3 - 11$ or $7{\cdot}3$ intervals of the decimal scale-height H' in the intervening 280 km. These $7{\cdot}3$ intervals H' are certainly not all equal; but their mean value is $38{\cdot}5$ km., which, for a mean molecular weight 20 (implying a considerable proportion of atomic oxygen in the upper levels) would correspond to a mean temperature, over this height interval, of just under $400°$ K. This seems not unreasonable at first sight, but it must be remembered that from 20 to 120 km. the average value of T seems unlikely to be less than $250°$; if the mean molecular weight there has a value about 27, this value of T would imply that over this height-interval, H' has the mean value $17{\cdot}4$ km.; consequently $100/17{\cdot}4$ or $5{\cdot}8$ of the $7{\cdot}3$ intervals H' would be included between 20 and 120 km., and n, at 120 km., would have the value $10^{18{\cdot}3 - 5{\cdot}8}$ or $10^{12{\cdot}5}$. The remaining $7{\cdot}3 - 5{\cdot}8$ or $1{\cdot}5$ intervals H', extending between 120 and 300 km., would have the high mean value $180/1{\cdot}5$ or 120 km.; this is too large (§ 5), and, taking the mean molecular weight there as 20, would imply a temperature of $1200°$ over this interval. Since the temperature at 120 km. may be less than $300°$, and no sudden transition there is to be expected, the mean value $1200°$ for T from 120 km. to 300 km.

would imply still greater temperatures in the upper part of this height-interval. Vegard [14.4 b], however, considers that the higher layers of the atmosphere extend far upwards on account of electrical forces, and that they are not at a high temperature.

The velocity of the ions will be smaller than that of the electrons in the ratio $\sqrt{(m_e/m_i)}$, which for an atomic oxygen ion is 1/172, and for a molecular nitrogen ion is 1/240. If the ionic mean free path is only about a quarter of that for an electron (corresponding to $d_i = 2d_e$), the collision interval for an atomic oxygen or molecular nitrogen ion will be about 43 ($= 172/4$) or 60 ($= 240/4$) times as long as that of an electron; thus τ_i/τ_e or ν_e/ν_i is approximately 50, or, if $d_e = d_i$, 200.

15.10. The spiral motion of charges in the upper atmosphere. A free electric charge in the presence of a uniform magnetic field of intensity H will execute a spiral motion round the lines of magnetic force. The projection of the path on a plane normal to the magnetic field is a circle, whose radius r may be called the *spiral radius* of the path. The angular velocity $\omega = v/r$ of this circular component of the motion is given by

$$\omega = eH/m, \tag{71}$$

e being the charge and m the mass of the particle; the spiral radius r is v/ω, where v is the component of the velocity of the particle, transverse to the field; v remains constant throughout the motion. The equation (71) expresses the balance between the centrifugal force mv^2/r or $m\omega^2 r$ and the deflecting force evH or $e\omega rH$ (since $v = \omega r$), Fig. 25. Another form of (71) is

$$Hr = mv/e. \tag{71 a}$$

In the earth's field we may take (1 (34))

$$H = H_0(1+3\cos^2\theta)^{\frac{1}{2}}, \tag{72}$$

where H_0 ($= 0\cdot3$ gauss) is the equatorial intensity, and θ is the geomagnetic co-latitude. Taking $m_i = 3\cdot3 \times 10^{-23}$ gm. for the ions (of molecular weight 20, say, intermediate between that for atomic oxygen and molecular nitrogen) we find, for ions and electrons respectively

$$\omega_i = 150(1+3\cos^2\theta)^{\frac{1}{2}}, \qquad \omega_e = 5\cdot3\times10^6(1+3\cos^2\theta)^{\frac{1}{2}}, \tag{73}$$

or, at the equator ($\theta = 90°$),

$$\omega_i = 150, \qquad \omega_e = 5\cdot3\times10^6.$$

An ion will therefore make ($\omega_i/2\pi =$) 25 rotations per second at the equator. If an average value for the transverse component of the electronic velocity be taken as 10^7, the electronic spiral radius r_e at the

equator is about 2 cm., and that of an ion (r_i) of molecular weight 20 will be greater in the ratio $\sqrt{(m_i/m_e)}$, or about 200 fold, giving a spiral radius of about 4 metres.

If the mean free paths, electronic (l_e) or ionic (l_i), are much smaller than these radii, the paths between collisions will be only slightly curved; but if l is comparable with, or greater than, the spiral radius, the spiral motion will have play, and will, for example, limit the rate of random travel or diffusion of electrons or ions transverse to the field, while the rate along the field will not be affected.
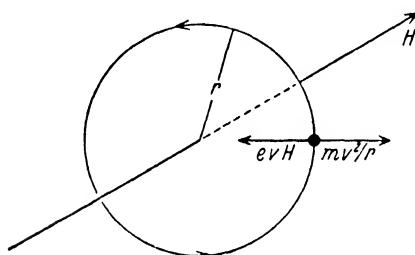

FIG. 25. The circular motion of a charged particle moving at right angles to a uniform magnetic field
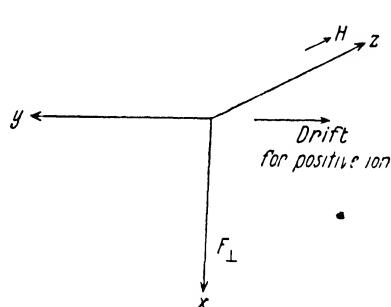
FIG. 26. The drift motion of a positive charge in a uniform magnetic field, due to a force perpendicular to this field

Suppose that, in addition to the electromagnetic force evH, some other force acts on the charged particle, with a component $F_{\parallel}$ along the magnetic field, and $F_{\perp}$ transverse to it. In this case the component motion along the magnetic field suffers an acceleration $F_{\parallel}/m$, as if that field were absent; the transverse force $F_{\perp}$, on the other hand, produces no mean velocity or acceleration in its own direction, but imparts a mean velocity or 'drift' to the charged particle, at right angles both to $F_{\perp}$ and the magnetic field, of amount $F_{\perp}/eH$, independent of the mass of the particle. If the z axis is taken along the direction of H, and the x axis along that of $F_{\perp}$, and if the remaining (y) axis is taken so that the x, y, z axes form a right-handed system, then the drift velocity is $-F_{\perp}/eH$ along the y axis. This readily follows from the solution of the equations of motion of the charge, which relative to these axes are

$$m\ddot{x} = F_{\perp} + eH\dot{y}, \tag{74}$$

$$m\ddot{y} = -eH\dot{x}, \tag{75}$$

$$m\ddot{z} = F_{\parallel}; \tag{76}$$

the solution for the velocities is

$$\dot{x} = \dot{x}_0 \cos \omega t + \dot{y}_0 \sin \omega t + (F_\perp/eH)\sin \omega t, \tag{77}$$

$$\dot{y} = -\dot{x}_0 \sin \omega t + \dot{y}_0 \cos \omega t - (F_\perp/eH)(1 - \cos \omega t), \tag{78}$$

$$\dot{z} = \dot{z}_0 + (F_\parallel/m)t, \tag{79}$$

where the suffix 0 refers to the initial values ($t = 0$). Thus the mean value of $\dot{x}$ is zero, that of $\dot{y}$ is $-F_\perp/eH$, and $\dot{z}$ is accelerated at the rate $F_\parallel/m$.

If the free charge forms part of a gas, and makes $\nu\,(= 1/\tau)$ collisions per second with other particles, after which its velocity takes a new direction at random, the accelerated motion *along* the magnetic field is constantly nullified by the collisions, and the force $F_\parallel$ merely produces a mean velocity $\frac{1}{2} F_\parallel \tau/m$ in its own direction. If the force $F_\parallel$ is due to an electric field E acting on the particle, then $F_\parallel = Ee$, the mean velocity is $\frac{1}{2}Ee\tau/m$, and the mean electric current-flow per particle is $\frac{1}{2}Ee^2\tau/m$; the corresponding current-intensity when there are n such charged particles present per unit volume is $\frac{1}{2}nEe^2\tau/m$. Dividing this by E, we obtain the electric conductivity κ of the gas, $\frac{1}{2}ne^2\tau/m$, or $ne^2\tau/m$ if we neglect the factor $\frac{1}{2}$, which we may do since the present calculation, and that of τ in § 11, neglects the statistical refinements of a proper treatment. If electrons and ions are both present, with number densities n_e and n_i, each set of particles makes its own contribution to κ, giving

$$\kappa_\parallel = \left(\frac{n_i \tau_i}{m_i} + \frac{n_e \tau_e}{m_e}\right)e^2 = \left(\frac{n_i}{m_i \nu_i} + \frac{n_e}{m_e \nu_e}\right)e^2; \tag{80}$$

since e is squared in these formulae, positive and negative ions (of like mass) make equal contributions to the electric conductivity, despite the difference of sign of their charges. This expression for $\kappa_\parallel$ (where the suffix $\parallel$ indicates that κ refers to current-flow parallel to the magnetic field H) does not involve H, implying that the electric conductivity in the direction of a magnetic field is unaffected by this field. In the absence of a magnetic field κ is the same for all directions.

The influence of a force $F_\perp$ transverse to the magnetic field is also affected by the collisions in a gas, which continually break off the motion represented by (77) and (78), and restart it with new values of $\dot{x}_0$ and $\dot{y}_0$. It is convenient to reckon the time afresh from zero at the instant of each collision, for the interval between that and the next collision, during which (77) and (78) apply. The average of $\dot{x}_0$ and $\dot{y}_0$ for a given charged particle, over a number of collisions, may be

assumed zero, so that the mean values of $\dot{x}$ and $\dot{y}$ in (77) and (78) depend only on the last term, involving $F_\perp$. But since the undisturbed motion represented by (77) and (78) does not continue indefinitely, the average value of $\sin \omega t$ and $\cos \omega t$ will not be zero, but will depend on ω and the average collision interval τ.

The actual collision intervals will take all values, with different probabilities: as is shown in books on the kinetic theory of gases, the fraction of the collision intervals which lie between t and $t+dt$ is $e^{-t/\tau} dt/\tau$. Thus the average value of $\sin \omega t$ in (77), at the end of the collision interval, is

$$\int_0^\infty e^{-t/\tau} \sin \omega t \, dt/\tau \qquad (81)$$

or

$$\int_0^\infty \nu e^{-\nu t} \sin \omega t \, dt, \qquad (82)$$

where ν is the collision frequency $(= 1/\tau)$. This integral is equal to

$$\frac{\nu \omega}{\nu^2 + \omega^2}. \qquad (83)$$

Hence the average $\dot{x}$ at the end of the collision interval is

$$\frac{F_\perp}{eH} \frac{\nu \omega}{\nu^2 + \omega^2} = \frac{F_\perp}{m} \tau \frac{\nu^2}{\nu^2 + \omega^2}. \qquad (84)$$

This is $\nu^2/(\nu^2 + \omega^2)$ times the z-velocity imparted at the end of the interval by an equal force F *along* the field. Thus the magnetic field does not now wholly destroy the tendency of a transverse force $F_\perp$ to produce an average flow of the charges in its own direction, but it reduces the flow in the ratio $\nu^2/(\nu^2 + \omega^2)$. Consequently, considering the case when $F_\perp$ is due to an electric field, we infer that the *transverse* electric conductivity $\kappa_\perp$, made up of ionic and electronic contributions, is given by

$$\kappa_\perp = \left\{ \frac{n_i \nu_i}{m_i(\nu_i^2 + \omega_i^2)} + \frac{n_e \nu_e}{m_e(\nu_e^2 + \omega_e^2)} \right\} e^2 = \left(\frac{n_i}{m_i \nu_i} \frac{r_i^2}{r_i^2 + l_i^2} + \frac{n_e}{m_e \nu_e} \frac{r_e^2}{r_e^2 + l_e^2} \right) e^2, \qquad (85)$$

where r is the spiral radius and l the mean free path. When the field is small, ω is small and r large, so that $\kappa_\perp$ approaches $\kappa_\parallel$, and becomes equal to it when $H = 0$. This reduction of the electric conductivity transverse to a magnetic field seems to have been first discussed theoretically by Gans; Pedersen [21] first called attention to its importance in the case of the atmosphere, in relation to the theory of the daily magnetic variations [23.14–18].

The average y component of velocity ($\dot{y}$ in (78)) at the end of a collision interval involves the average value of $\cos \omega t$, and the integral

$$\int_0^\infty e^{-t/\tau} \cos \omega t \, dt/\tau \tag{86}$$

or

$$\int_0^\infty \nu e^{-\nu t} \cos \omega t \, dt, \tag{87}$$

which has the value

$$\frac{\nu^2}{\nu^2 + \omega^2}. \tag{88}$$

Hence the average value of $1 - \cos \omega t$ is $\omega^2/(\nu^2 + \omega^2)$, and the mean value of $\dot{y}$ at the end of a collision interval will be

$$-\frac{\omega^2}{\nu^2 + \omega^2} \frac{F_\perp}{eH} \quad \text{or} \quad -\frac{l^2}{r^2 + l^2} \frac{F_\perp}{eH}. \tag{89}$$

When the gas is relatively dense, so that l is small compared with r, the drift velocity has only a small fraction of its full value, while under these conditions $\kappa_\perp$ differs little from $\kappa_\parallel$. When the gas is relatively rare, and l is large compared with r, the drift velocity has nearly its full value, and $\kappa_\perp$ is much less than $\kappa_\parallel$.

15.11. The electrical conductivity of the upper atmosphere transverse to the magnetic field. The electrical conductivity κ in the absence of a magnetic field is the same as the 'longitudinal' conductivity $\kappa_\parallel$ along a magnetic field if present. The contribution made to $\kappa_\parallel$ by *each* charge e of mass m is

$$\frac{e^2}{m\nu} \quad \text{or} \quad \frac{e^2}{m}\tau; \tag{90}$$

this increases indefinitely with increasing height. The complete expression for $\kappa_\parallel$ is

$$\left(\frac{n_i}{m_i \nu_i} + \frac{n_e}{m_e \nu_e} \right) e^2; \tag{91}$$

at very great heights the air must become almost completely ionized, and there $n_i = n_e = \frac{1}{2}n$, where n is the total number-density; since ν is proportional to n, it appears that $\kappa_\parallel$ tends to a finite limit at very great heights.

The transverse electric conductivity $\kappa_\perp$, for current-flow at right angles to the magnetic field, receives a contribution from each charge per unit volume, of amount

$$\frac{e^2}{m} \frac{\nu}{\nu^2 + \omega^2}. \tag{92}$$

This expression tends to zero both when ν tends to infinity and to zero, that is, at very great and very low densities; it has a maximum value $(\sqrt{e^2})/2H$ when $\nu^2 = \omega^2$, where $\sqrt{e^2}$ is written instead of e, because the expression is to be positive whatever the sign of e. This maximum value is independent of the mass and sign of the charge, and is therefore the same for ions and electrons. Taking $H = 0.4$ gauss as an average value, it is

$$2 \times 10^{-20}\,\text{e.m.u.} \tag{93}$$

At other levels the $\kappa_\perp$-contribution per charge is

$$\frac{2\nu\omega}{\nu^2+\omega^2} \tag{94}$$

times this maximum value. When ν/ω or ω/ν is 3, this fraction is 0·6, and for ν/ω or ω/ν equal to or greater than 10, the fraction is approximately equal to the lesser of $2\nu/\omega$ or $2\omega/\nu$ (e.g. it is 0·2 if ν/ω or ω/ν is 10). Since ω is a constant for any given point on the earth (except for the slow variation of H with height), and ν varies by a factor 10 in a range of height H' (§ 2)—supposing T constant— it follows that the region within which a charge of any kind can contribute, to $\kappa_\perp$, an appreciable fraction of the maximum possible amount, given by (93), is confined to a layer of thickness $2H'$ centred at the level at which $\nu = \omega$.

This level is very different for electrons and ions. If $H = 0.4$ gauss, and the ions are of molecular weight 20, then $\omega_e = 7 \times 10^6$, $\omega_i = 2 \times 10^2$, and ν_e/ν_i at any level is 50, or perhaps 200 (§ 9). If the 'observed' values of ν_e given in § 9 for the F_1 and F_2 layers are correct, namely 3.6×10^3 and 1.6×10^3, these layers are much above the level at which $\nu_e = \omega_e$, and each electron there contributes to $\kappa_\perp$ only a fraction $\frac{1}{1,000}$ and $\frac{1}{2,200}$ of its possible maximum given by (93). Taking the F_1 and F_2 values of ν_i as 72 and 32, the ions at those levels contribute fractions $\frac{2}{3}$ and $\frac{1}{3}$ of their possible maximum. The full contribution per ion will be made at a depth approximately $\frac{1}{2}H'$ below the level in the F_1 layer to which the measurement $\nu_e = 3.6 \times 10^3$ refers. Since $n_i \geqslant n_e$, it follows that the $\kappa_\perp$ for the F_1 and F_2 layers must be almost wholly due to the ions there. If n_i were independent of height, then (taking

$$\nu_i = \omega_i 10^{-h/H'} = \omega_i e^{-2\cdot 3h/H'} = \omega_i x,$$

where h here denotes height measured from the level at which $\nu_i = \omega_i$) the integral of the ionic part of $\kappa_\perp$ over all heights would be less than

$$\frac{n_i e^2}{m_i} \int_{-\infty}^{\infty} \frac{\nu_i}{\nu_i^2+\omega_i^2}\,dh = \frac{n_i e}{H}\frac{H'}{2\cdot 3}\int_0^{\infty}\frac{dx}{1+x^2} = \frac{\pi}{2}\frac{n_i e H'}{2\cdot 3H}\,, \tag{95}$$

the same as the integral of the ionic $\kappa_\perp$ for a layer of *uniform* density, in which $\nu_i = \omega_i$ throughout, and which is of thickness $\pi H'/2\cdot3$. Hence the total ionic contribution to the height-integral of $\kappa_\perp$ throughout the F_1 and F_2 layers must be less than

$$\tfrac{1}{2}\pi(n_i)_{\max} eH'/2\cdot3H, \tag{96}$$

where $(n_i)_{\max}$ is the maximum value of n_i in these layers. If this is taken to be the same as the maximum value of n_e, or say 10^6, and if $H'/2\cdot3$ is taken to be 50 km., the resulting value of the integrated ionic transverse conductivity (and therefore of the total transverse conductivity) in the F layers will necessarily be less than

$$3\times10^{-7} \text{ e.m.u.} = 300 \text{ ohm}^{-1}. \tag{97}$$

Four-fifths of the expression (95) is contributed by the range of x which lies between $x = \tan(\pi/20)$ and $x = \cot(\pi/20)$, or $x = 1/6$ and $x = 6$; this corresponds to a layer of thickness $2H'\log_{10}6$ or roughly $1\cdot6H'$, half above and half below the level at which $\nu_i = \omega_i$. The decimal scale-height H' of the F region is $2\cdot3$ times the ordinary scale-height (say 50 km., § 5), so that $H' = 115$ km.; $\nu_i = \omega_i$ at a level about $\tfrac{1}{2}H'$ below the F_1 height at which $\nu_e = 3\cdot6\times10^3$ (say 265 km.), that is, at about 210 km.; this level is therefore situated well above the E layer. If the value of n_i were independent of the height, four-fifths of the ionic part of $K_\perp$, the height integral of $\kappa_\perp$, would be contributed by the layer between the heights $210\pm(0\cdot8H')$ km.; if H' were everywhere 115 km., this would be between 120 and 300 km., but below 150 km. this value of H' is certainly excessive, so that the lower limiting height just stated should be raised perhaps to 140 or 150 km. Thus if $(n_i)_{\max}$ were 10^6, the ionic $K_\perp$ would be mainly contributed by the top of the E layer, and the F_1 layer. But in the E layer $n_i = n_+ + n_- = 2\lambda n_e$ approximately (p. 526), and the maximum value of this is $4\lambda \times 10^5$, taking $n_e = 2\times10^5$, $\lambda = 100$; in the F_2 layer, where λ is of order 1 or less, and n_e has the maximum value 10^6, the maximum value of n_i will likewise be about 10^6. Hence, if $\lambda = 100$ in the E layer, n_i has a maximum 40 times greater in the E than in the F layer; and in the E layer the ratio $2\nu_i\omega_i/(\nu_i^2+\omega_i^2)$ is probably nearer unity than in the F_2 layer. Hence the ionic part of the height integral of $K_\perp$ is likely to be due mainly to the E layer, and to be of the order 12×10^{-6} e.m.u. (though definitely less than this amount).

The expression (96), with n_e substituted for n_i, gives the maximum possible contribution of the electrons to the transverse conductivity. The greatest maximum of n_e is 10^6, and occurs in the F_2 layer; but there

the factor $2\nu_e\,\omega_e/(\nu_e^2+\omega_e^2)$ is very small ($<1/2000$), so that the electronic contribution made by the F layer to the integral transverse conductivity is less than one-thousandth of the value (97), that is, less than 3×10^{-10} e.m.u.

In the E layer the factor $2\nu_e\,\omega_e/(\nu_e^2+\omega_e^2)$ for the electrons is greater than in the F layers, but even there, if the value of ν_e is about 10^4 (p. 501), this factor is no greater than $1/300$; the maximum value of n_e in the E layer is only 2×10^5, and H' is less there than in the F layers. Hence the contribution of the electrons in the E layer to $K_\perp$ is less than 10^{-10} e.m.u. Not till we attain the level where $\nu_e=7\times10^6=\omega_e$ does the ratio $2\nu_e\,\omega_e/(\nu_e^2+\omega_e^2)$ for the electrons attain its maximum value 1; this level is about $2\cdot8H'$ below 120 km., the height of the E layer maximum, at which $\nu_e=10^4$. Taking H' there as equal to 25 km., this level is about 50 km.; and at that height the value of n_e must be relatively small, except during radio fade-outs.

15.12. Atmospheric motions. A property of the upper atmosphere which is of importance for geomagnetism is the wind, regular or irregular, in the ionosphere. Meteor trails [57, 58] occasionally reveal the existence of strong winds in the lower part of the ionosphere: sometimes the trails are twisted, suggesting local irregularities of wind direction, or turbulence. Volcanic dust also gives evidence of strong winds at high levels; by their agency it is spread over the earth from the region of eruption, as is shown especially by abnormal sunrise and sunset glows after great eruptions. The high luminous night clouds, observed and measured by Jesse and Störmer, also show considerable motion [49–51]. But the motions which appear to be of most importance for the geomagnetic variations are the regular daily periodic circulations [64–9] which probably extend throughout all atmospheric levels, and manifest themselves at the ground by producing daily variations of the barometric pressure. In the tropics there is a small but clearly observable solar semi-diurnal variation of the pressure; this can be determined also in the records of stations in middle and high latitudes, but only by averaging the records from many days, since there the small regular variation is overlaid by much larger changes associated with weather—cyclones and anticyclones. The maximum pressure associated with the variation occurs at about 10 a.m. and p.m. of local time. The solar semi-diurnal oscillation can also be detected in the wind observations, but the air currents associated with it, near ground-level, are quite small (about 20 cm./sec. or $\frac{1}{2}$ mile/hour) compared with the ordinary systematic or irregularly variable winds.

Besides the solar semi-diurnal variation of the pressure and the wind, there are smaller but fairly regularly distributed 8- and 6-hourly variations, and a less regularly distributed 24-hourly variation, relatively large at some stations, and small over the oceans.

A still smaller barometric variation, but one of great theoretical interest, is caused by the moon. It is in fact a lunar atmospheric tidal variation, and is purely semi-diurnal. It is nearly 40 per cent. greater at perigee than at apogee, like the lunar tide-producing force. Its geographical distribution is very similar to that of the solar semi-diurnal barometric variation, but its amplitude, and that of the small winds associated with it, are only about one-tenth of those of the solar variation. On account of its small magnitude, it can be determined accurately only by averaging the data over several years.

Fig. 27 illustrates [67] the distribution of the annual mean lunar atmospheric tide, thus found, at a number of stations. The arrows on the map are centred at the stations named, and their lengths are proportional, on the scale shown, to the maximum lunar tidal departure of the barometric pressure from its mean value, at each station. The direction of the arrows, if these are interpreted as hour hands on a clock keeping lunar time (so that the upward vertical corresponds to the upper or lower passage of the moon across the meridian) indicates the time of high pressure due to the tide. (Note that the arrow for Melbourne is incorrect; it should be turned through 30° towards the vertical.) The times of high tide agree with the times of lunar passage across the meridian, to within an hour or two, in the average of the year: at most stations high tide occurs rather less than an hour after lunar transit.

This is not so, however, in the months of December and January, when the high tide is further retarded by about an hour, at nearly all stations. This occurs simultaneously at stations north and south of the equator, so that though it is an annual change of phase it is not an ordinary *seasonal* change, due to the passage of the sun north or south of the equator.

The change of phase (and amplitude) of the tide at Hong Kong, from month to month throughout the year, is illustrated [67] in Fig. 28, where the point shown for each month is to be interpreted as the end of an arrow drawn from the centre of the outer quadrant, this arrow representing the amplitude and the time of high tide, as in Fig. 27.

The annual mean phase of the lunar atmospheric tide is roughly such
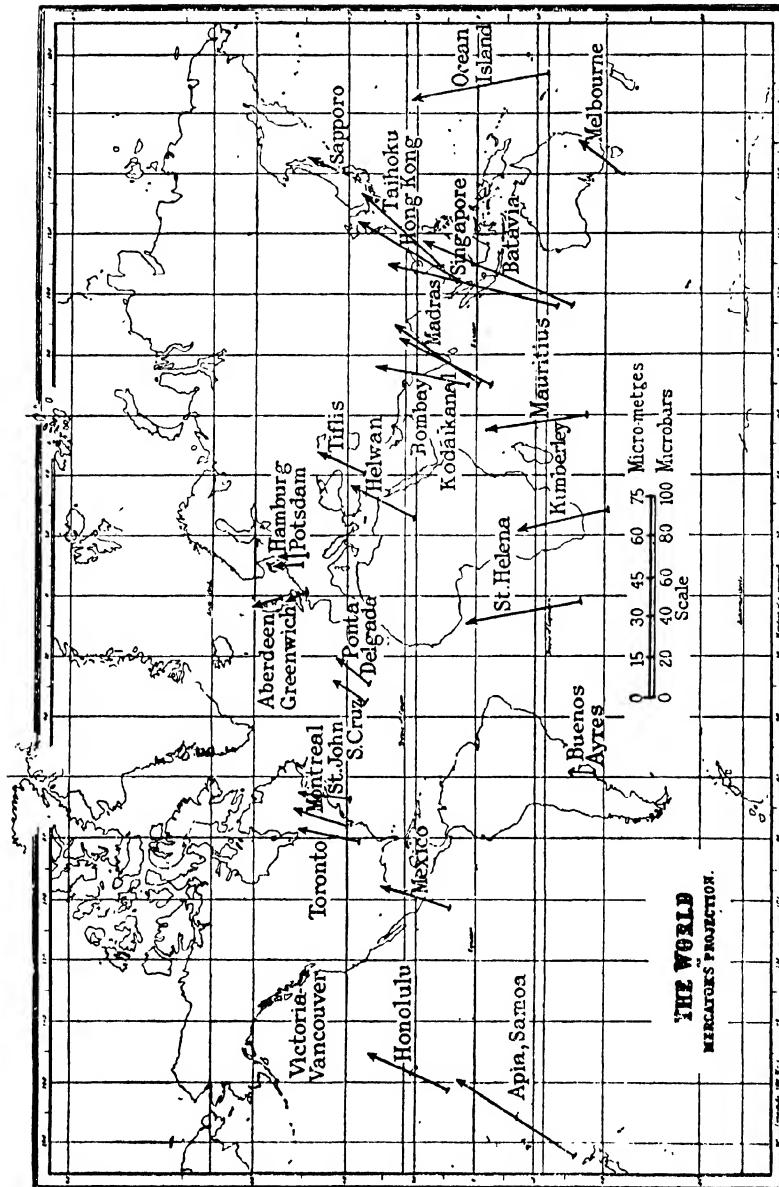
Fig. 27. Map showing the world-distribution of the lunar atmospheric tide. Each arrow or tide-vector indicates the amplitude of the lunar tidal variation of the barometric pressure (on the given scale) at the terrestrial point at the centre of the arrow. The direction of the arrow indicates the local lunar time of maximum average barometric pressure, as shown on the dial of a clock keeping lunar time (12^h representing the time when the moon crosses the meridian)

as a simple tidal theory would suggest: its magnitude also is not much larger than such a theory would indicate. But the solar semi-diurnal barometric variation is mysterious both in amplitude and phase: if it, too, were purely tidal, its maximum should occur at or after 12 noon and midnight, instead of at 10 a.m. and 10 p.m.; and its amplitude would be expected to be smaller than that of the lunar tide, in the ratio 2/5
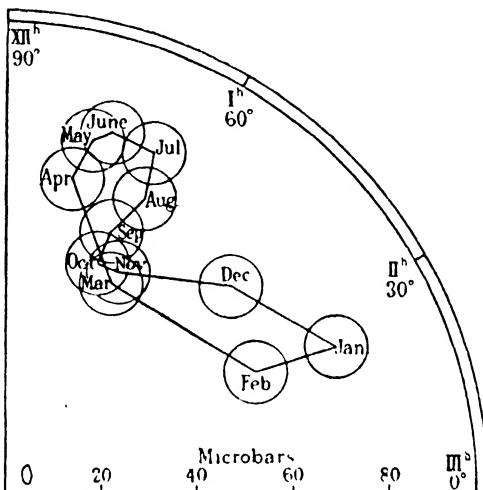

FIG. 28. Harmonic dial indicating the variation, throughout the year, of the amplitude and phase (or time of maximum) of the lunar tidal variation of barometric pressure at Hong Kong. The centre of each circle (a probable-error circle) represents the end of the tide-vector (drawn from the lower left-hand corner as origin) for the month indicated

(that of the solar to the lunar tidal force) instead of being fifteen times as great. The large magnitude of the solar semi-diurnal variation seems to be due to magnification by resonance, the period of 12 solar hours happening to coincide within a few minutes with a natural period of oscillation of the atmosphere, of corresponding type. This magnifies the solar atmospheric tide by a large factor (of the order 100), and likewise magnifies an otherwise small semi-diurnal variation of thermal origin; the two parts, tidal and thermal, combine into one variation, with the observed phase. The development of this hypothesis, begun by Kelvin, has recently been much advanced by Taylor [64] and Pekeris [65], whose work suggests that the oscillation is reversed in phase at great heights, and also is much magnified. Consequently, the associated wind, of speed about $\frac{1}{2}$ mile per hour near the ground, may be 50 miles

per hour in the ionosphere, and therefore probably comparable with other less regular winds that may blow there.

If the solar semi-diurnal circulation undergoes such a reversal and magnification in the high atmosphere, it is not unlikely that so also does the lunar atmospheric tide. But these statements about either the solar or the lunar semi-diurnal circulations at great heights are necessarily uncertain, since (i) the numerical theory of them involves some uncertain assumptions, (ii) the lunar atmospheric tide undergoes a considerable annual variation of unknown cause, and (iii) large independent solar daily periodic motions may be generated in the upper ionosphere, if great daily temperature variations occur there.

Appleton and Weekes [70] have inferred, from direct reflection observations of radio waves, a lunar semi-diurnal variation of the equivalent height of the region E of the ionosphere. The motion is tidal in character; the maxima occur about $\frac{3}{4}$ hour before the upper and lower lunar transits. The total range is about 1·7 km. These observations suggest a pressure oscillation of relative amplitude $(\Delta p/p)$ 0·068 at a level of 120 km., whereas the measurement of the lunar barometric oscillations at ground-level shows a corresponding relative amplitude of 0·0000115. The proportionate lunar tidal change of pressure at 120 km. thus appears to be nearly 6,000 times as large as that at ground-level.